AWARD NUMBER:

W81XWH-10-1-0122

TITLE:

RNA-Guided DNA Rearrangements in Breast Cancer

PRINCIPAL INVESTIGATOR:

Wenwen Fang, Ph.D .

CONTRACTING ORGANIZATION:

Princeton University, Princeton, NJ, 08544

REPORT DATE:

September 2012

TYPE OF REPORT:

Annual Summary

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

DISTRIBUTION STATEMENT: Approved for Public Release;
Distribution Unlimited

The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision unless so designated by other documentation.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. **PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.**

1. REPORT DATE *(DD-MM-YYYY)*	**2. REPORT TYPE**	**3. DATES COVERED** *(From - To)*
1 September 2012	Annual Summary	1 Mar 2010 – 30 Sep 2012

4. TITLE AND SUBTITLE
RNA-Guided DNA Rearrangements in Breast Cancer

5a. CONTRACT NUMBER
W

5b. GRANT NUMBER
W81XWH-10-1-0122

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Wenwen Fang
fang.wenwen@gmail.com

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) AND ADDRESS(ES)
Princeton University, Princeton, New Jersey, 08544

8. PERFORMING ORGANIZATION REPORT

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel
Command, Fort Detrick, Maryland 21702-5012

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

Genome rearrangement and instability is a hallmark of cancer. In light of a previous study from our lab demonstrating RNA-templated DNA rearrangements in *Oxytricha,* I searched for chimeric transcripts in normal human mammary cells to address the hypothesis that such chimeric RNAs may occasionally guide DNA rearrangements during breast tumorigenesis. By both computational and experimental analyses, I showed that even normal human cells produce chimeric RNAs, likely via RNA *trans*-splicing without corresponding DNA rearrangement. The fact that rearrangements at the level of RNA can precede that of DNA suggests the possibility that the presence of chimeric RNA may predispose the DNA genome to rearrangements.

In addition, I utilize the programmed genome remodeling in the ciliate *Oxytricha* as a model system to understand how a class of small RNAs called piRNAs facilitates genome-wide rearrangements. Through a combination of molecular, high-throughput sequencing, and synthetic biology approaches, I provide evidence for a model where piRNAs protect DNA against loss during *Oxytricha* genome rearrangement. This not only reveals a novel function for piRNAs, but also underscores a plasticity of RNA-based regulatory systems, because in two distantly-related ciliate species, small RNAs target DNA for deletion instead during genome reduction.

15. SUBJECT TERMS
Transcriptome, RNA trans-splicing, genome instability, Piwi-interacting RNA, genome rearrangement

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
					USAMRMC
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**	UU	63	**19b. TELEPHONE NUMBER** *(include area code)*
U	U	U			

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. Z39.18

Table of Contents

Introduction

Genome instability plays a pivotal role in the evolution of breast cancer cells, granting them an intrinsic ability to gain cancer-associated phenotypes and to develop resistance to targeted therapies [1-3]. Facilitated by high throughput sequencing techniques, breast cancer genome and transcriptome sequencing data are now accumulating at an ever-growing speed, and this provides us with a starting point to understand associated chromosomal rearrangements and their role in tumor formation and progression [4-9]. Inspired by the novel finding in my co-mentor's lab that RNA can guide genome rearrangements in the ciliated protozoan *Oxytricha* [10], I decided to investigate RNA's role in DNA rearrangements associated with breast cancer. Specifically, I tested 1) whether RNA can guide corresponding DNA rearrangement in breast cancer cells, 2) whether normal breast tissue and cells contain abnormal transcripts that may act as natural templates for DNA rearrangements, and 3) what function the small non-coding RNAs carries out in *Oxytricha* genome rearrangement.

Body

The following section is written according to my approved Statement of Work (SOW).

Aim 1. Test RNA-guided DNA rearrangement in breast cancer cells.

The overall design for this aim is to deliver an RNA template into breast cells and test whether the supply of RNA templates promotes corresponding DNA rearrangement. During the first year of my funding period (March 2010 ~ March 2011), I tested whether expression of an RNA template would encourage the corresponding deletion in *BRCA2* (Figure 1A , 2011 annual report). This intragenic deletion originally occurred in CAPAN1 pancreatic cancer cell lines after DNA-damaging drug treatment[11], and the presence of short homology at the deletion boundary resembles the short direct repeats at DNA recombination junctions in *Oxytricha* genome rearrangement. I first attempted the test in CAPAN1 cells by transient transfection of a plasmid expressing the RNA template. Surprisingly, the deletion was detected even in parental non-transfected cells (Figure 1B, 2011 Annual report), making the results difficult to interpret. I then established stable lines expressing the same RNA templates, but screening all these stable lines did not suggest any evidence of RNA-templated DNA recombination (Figure 1C , 2011 annual report). I also delivered RNA templates to MDA-MB-231 cells, which are breast cancer cells, but could not detect any evidence of RNA-guide DNA recombination, with or without DNA damaging drug treatment (Figure 2 and 3, 2011 annual report). It is likely that the sensitivity of PCR assays, and even nested PCR assays, are not sufficient for our purpose. For detailed description of this part, see my 2010 biannual report, 2011 annual report and its revised version.

To increase the detection sensitivity, I modified the experimental design. Under the revised statement of work, I delivered a linear luciferase expression plasmid and an RNA template to breast cancer cells MDA-MB-231, and tested whether the RNA template can facilitate re-circularization of the luciferase plasmid, leading to its expression which can be detected by luciferase-based assays. I used PCR to produce linearized luciferase plasmid, and used *in vitro* transcription to prepare RNA template. I then co-transfected the linear DNA substrate and RNA template, by using TransIT-mRNA kit (Mirus). Single transfection of either the linear DNA substrate or the template, or water transfection served as negative controls, and

transfection of circular DNA plasmid served as a positive control. I did detect a signal from the 2-day positive control-transfection sample, but all the other transfections show background level as in the negative control (Figure 1D, 2012 annual report). Therefore, I was not able to conclude that the expected re-circularization occurred in the experiments. For details of this section, refer to 2011 biannual report and 2012 annual report.

In summary, my direct experimental test for RNA-guided DNA rearrangement in breast cells did not yield positive result. This does not mean that such a phenomenon does not exist. In fact, RNA-templated DNA repair, originally reported in yeast[12], was recently extended to bacteria and human cells[13], providing a strong support for the possibility of RNA-guided DNA rearrangement in cancer. I anticipate that further studies utilizing different systems, such as recurrent genome rearrangement in cancer, will be very valuable in testing RNA-guided DNA rearrangement.

Aim 2. Characterize RNA fusions in normal breast cells and breast cancer cells. (Months 18 ~ 36)

In this aim, I proposed to collect RNA-seq data of normal breast cells and breast cancer cells and computationally search for fusion transcripts. I first used customized script to search for RNA fusions involving *BCAS3*, a gene highly expressed in breast cancer[14] and commonly fused to *BCAS4*[8]. I detected 10 intergenic chimeric transcripts involving *BCAS3* in normal human epithelial keratinocytes (NHEK) using ENCODE RNA-seq data, and found an AGG/CCT motif at the fusion junction, suggesting that these transcripts likely derived from *trans*-splicing (Figure 4 and 5, revised 2011 annual report). Because the customized program can only search fusions involving one candidate gene at a time, I later implemented deFuse[15], a published program to search genome-wide RNA fusions systematically. Using this program, I confirmed 13 previously reported chimeric RNAs, and discovered 5 new RNA fusions in breast cancer cells (Table 3, revised 2011 annual report). I also detected two intragenic deletions in normal breast cells (Figure 6, revised 2011 annual report), which supports our hypothesis that even normal cells produce "abnormal" RNAs that may potentially influence genome rearrangement.

The release of high-throuput RNA sequencing data for Human Mammary Epithelial Cells (HMEC, normal karyotype) and MCF7 cell from the ENCODE project provided me with ideal datasets to characterize chimeric RNAs in normal mammary and breast cancer cells. In the meantime, the deFuse program was updated to allow users to retrieve reads that correspond to each fusion prediction. I was able then to further characterize predicted fusions and eliminate false positives. Because of computation memory problem I had to divide RNA-seq datasets into subsets (20 million paired-end reads for HMEC, and 1 million paired-end reads for MCF7), which may lower the detection sensitivity. Nevertheless, I predicted 95 unique RNA fusions in the HMEC dataset, 38 of which were found in two or more subsets; 209 fusions in MCF7 dataset, 17 of which were identified independently. I also predicted 120 unique RNA fusions in the 10 million read data for NHEK (Table 1, 2012 annual report). The majority of predicted RNA fusions in HMEC cells are intra-chromosomal and adjacent, derived from alternative splicing or read-through (Table 2, 2012 annual report). In contrast, the predicted RNA fusions in MCF7 and NHEK cells are mostly inter-chromosomal and not adjacent, not derived from alternative splicing or read-through. The majority of the fusions do not preserve open reading

frame in all three datasets. Among those intra-chromosomal fusions, I found two fusions, *TANC2-MIR633* and *ZNF782-ZNF510*, that are recurrent in HMEC and MCF7 cells; two fusion, *EHD2-GLTSCR2* and *PLCH2-RP4-740C4*.4, that are shared by HMEC and NHEK cells (Table 1, 2012 annual report).

Among all the predicted RNA fusions, interchromosomal chimeric RNAs are the most interesting, because they are thought to be rare in cells with normal karyotype (no genomic rearrangements) and may influence long-range chromosomal rearrangements. From the HMEC RNA-seq data, I predicted ten inter-chromosomal RNA fusions: *ARR3-SNORD56B*, *GAPDH-PDCD4*, *KRT5-KRT14*, *KRT6A-KRT14, FAM62B-XPNPEP3, TUBA4A-TUBA1A, TUBA4A-TUBA1C, TUBB2B-TUBB2C, TUBB6-TUBB2C*, and *ZC3HAV1L-CHMP1A* (Table 1, 2012 annual report). I then focuses on one of them, *ZC3HAV1L-CHMP1A*, for experimental verification. I chose this one because of several reasons. First, this prediction was found independently in two subsets of data, making it a more confident one. Second. The fusion has an over 0.9 probability score, again suggesting an authentic fusion. Third, the fusion preserves the open reading frame, leading to a possibility of a novel protein derived from the fusion transcript. The experimental characterization of this chimeric RNA is detailed in 2012 annual report, and the submitted manuscript "Detection of a common chimeric transcript between human chromosomes 7 and 16" (see appendices). Briefly, my RT-PCR analysis confirmed the presence of *ZC3HAV1L-CHMP1A* RNA fusion in HMLE cells, a human mammary cell line derived from HMEC, as well as a panel of other human cell lines and tissues. I also recovered some variants of this chimeric transcripts, though at much lower levels. Because the RNA fusion, including its variants, follows canonical splicing signal, it likely derives from *trans*-splicing, rather than reverse transcription artifact. Finally, I was able to detect the entire predicted open reading frame for the chimeric transcript by nested PCR, suggesting that it may be translated into a novel fusion protein (Figure 1, manuscript 1 in appendices). Please also refer to attached manuscript submitted to *Biology Direct*.

To summarize, I have reached my goal of characterizing RNA fusions both computationally and experimentally, especially in normal mammary cells, because such chimeric transcripts may occasionally impact genome rearrangement according to our hypothesis. The apparent splicing signal at the fusion junction suggests that *trans*-splicing is a major mechanism for generating chimeric transcripts, and it occurs more often in normal cells than previously thought.

Aim 3. Test the function of small RNAs in *Oxytricha* genome rearrangement.

The unique biology of dual genomes in the ciliated protozoan *Oxytricha* provides an excellent model system to study genome rearrangement. The genome remodeling not only occurs extensively, but also in a developmentally regulated fashion. Previously studies utilizing this system already demonstrated a novel function for long non-coding RNAs to mediate genome remodeling in *Oxytricha*[10]. In this aim, I focused on a class of ~27 nt small RNAs, which also accumulate to very high levels during genome rearrangement in *Oxytricha*.

I first showed that Otiwi1, a Piwi protein in *Oxytricha* co-expresses and interacts with the 27 nt small RNAs (thus called piRNAs for Piwi-interacting RNAs) (Figure 1, manuscript 2 in appendices), and its localization shifts from the parental somatic nucleus to the developing

nucleus where genome rearrangement occurs (Figurer 2, manuscript 2 in appendices). Knockdown of Otiwi1 leads to the depletion of the 27 nt piRNAs, as well as a developmental arrest and disintegration of *Oxytricha*, suggesting that Otiwi1 is essential for the developmental stage when genome rearrangement occurs (Figure 3, manuscript 2 in appendices). Deep-sequencing analyses suggest the Otiwi1-bound piRNAs derive from the entire somatic genome of *Oxytricha*, instead of the germline genome as in other distantly related species of ciliates (Figure 4, manuscript 2 in appendices). This led to our hypothesis that the 27 nt piRNAs in Oxytricha specify genomic regions for retention during the rearrangement process, because sequences not in the somatic genome are discarded from the germline genome during development. Finally, injection of synthetic RNAs mimicking endogenous piRNAs but correspond to normally deleted regions leads to their retention (Figure 5, manuscript 2 in appendices), which provides a direct support for our model that piRNAs protect DNA against loss in *Oxytricha*. For details of these experiments, please see 2011 and 2012 annual report, and the attached manuscript accepted by *Cell*.

Our novel finding that piRNA protects DNA against loss during *Oxytricha* genome rearrangement is of great interest and importance to the fields of RNA and genome biology, and was well received in multiple conferences where I presented the work. Recent studies of small RNAs in the nematode *C. elegans* also suggests a subset of piRNAs may specify "self", making an interesting parallel to our work. Although this novel function of piRNA has not been extended to human cells, they will definitely inspire new directions in cancer biology. Piwi proteins, for example, have already been shown to regulate tumorigenesis[16].

Key Research Accomplishments

- Computationally predicted an inter-chromosomal RNA chimera *ZC3HAV1L-CHMP1A* in HMEC cells.
- Experimentally verified that the *ZC3HAV1L-CHMP1A* fusion exists in normal beast cells including HMEC and MCF10A, and a breast cancer cell line MB-MDA231. It is also present in a panel of human tissue types, suggesting that it is a common RNA fusion in human.
- Demonstrated that Piwi-interacting RNAs protect DNA against deletion during *Oxytricha* genome rearrangement.

Reportable Outcomes

- Ph.D. degree obtained at the Department of Molecular Biology, Princeton University
- Poster presentation at the Era of Hope Breast Cancer Research Program Conference (Aug. 2011)
    - Title: Testing RNA-guided DNA rearrangements in breast cancer
- Oral presentation at Cell Symposium on Regulatory RNAs (Oct. 2011)
    - Title: Piwi-interacting RNAs protect DNA against loss during *Oxytricha* genome rearrangement
- Oral presentation at the Genome Integrity Discussion Group meeting (Dec. 2011)
    - Title: Piwi-interacting RNAs protect DNA against loss during *Oxytricha* genome

rearrangement
- Oral presentation at the Cold Spring Harbor Biology of Genome meeting (May, 2012)
    - Title: Piwi-interacting RNAs protect DNA against loss during *Oxytricha* genome rearrangement
- Manuscript titled "Piwi-interacting RNAs protect DNA against loss during *Oxytricha* genome rearrangement" accepted by *Cell*
- Manuscript titled "Detection of a common chimeric transcript between human chromosomes 7 and 16" submitted to *Biology Direct*
- Manuscript titled "Genomes on the edge: Programmed Genome Instability in Ciliates" submitted to *Cell*

Bibliography of all publications supported by this grant

Fang W., Wei Y., Kang Y., Landweber L. F. Detection of a common chimeric transcript between human chromosomes 7 and 16. *Submitted*
Bracht J. R., **Fang W.**, Stein E. M., Goldman A. D., Dolzhenko E., Landweber L. F. Genomes on the edge: Programmed Genome Instability in Ciliates. *Submitted*
Fang W., Wang X., Bracht J., Nowacki M., Landweber L. F. Piwi-interacting RNAs protect DNA against loss during *Oxytricha* genome rearrangement. *Cell, accepted*
Fang W., Landweber, L. F. RNA-mediated genome rearrangement: Hypotheses and evidence. *In preparation*

Meeting abstracts

For *Era of Hope Breast Cancer Research Program Conference:*

Genome instability is an early step in breast cancer tumorigenesis. Despite a more comprehensive sequence analysis of cancer genomes, our understanding of the initiation of genome rearrangement in these cells is far from complete. The novel cellular activity of RNA-guided DNA rearrangement, originally discovered in *Oxytricha* in the Landweber laboratory, revealed a hidden layer of genome regulation that may also contribute to cancer. In this study, we aim to test whether RNA-guided DNA rearrangement also exists in breast cancer cells and to gain a better understanding of the process of RNA-guided DNA rearrangement.

In CAPAN-1 cancer cell lines, large intragenic deletions have been reported. We transfected these cells with a construct expressing an RNA template that mimics such deletions. Using PCR assay, we detected corresponding deletion in the genomic DNA of transfected cells. However, further experiments show that they may occur spontaneously in natural population as well. We are testing whether similar phenomenon exists in breast cancer cell lines. Additionally, we are searching for fusion transcripts that might be natural RNA templates, by analyzing deep-sequencing data.

For *Cell symposium on Regulatory RNAs, Genome Integrity Discussion Group Meeting, Cold Spring Harbor Biology of Genome Meeting:*

Genome duality in ciliated protozoa offers a unique system to showcase their epigenome as a model of inheritance. In *Oxytricha,* the somatic genome is responsible for vegetative growth, while the germline contributes DNA to the next sexual generation. Somatic nuclear development

eliminates all transposons and other so-called "junk DNA", which constitute ~95% of the germline. We demonstrate that Piwi-interacting small RNAs (piRNAs) from the maternal nucleus can specify genomic regions for retention in this process. *Oxytricha* piRNAs map primarily to the somatic genome, representing the ~5% of the germline that is retained. Furthermore, injection of synthetic piRNAs corresponding to normally-deleted regions leads to their retention in subsequent generations. Our findings highlight small RNAs (sRNAs) as powerful transgenerational carriers of epigenetic information for genome programming.

Conclusion

Although direct testing of RNA-guided DNA rearrangements have not yielded promising results, I have demonstrated successful computational search for chimeric RNAs in normal keratinocytes, normal breast cells and breast cancer cells utilizing high-throughput RNA-seq data that are readily available. I experimentally verified one inter-chromosomal RNA fusion (*ZC3HAV1L-CHMP1A*) that is predicted from normal mammary cells, and found that it is also recurrent in many other human tissue types. Furthermore, I found three splicing variants of this chimeric RNA, suggesting promiscuous splicing involving the two RNA transcripts.

Combining deep-sequencing, loss-of-function study, and "gain-of-function" studies in *Oxytricha*, we proposed a model where Piwi-interacting RNAs protect genomic regions for retention during genome rearrangements. Although whether this phenomenon extends to human cells needs further test, the fact that Piwi proteins show elevated expression hints a possible link between piRNA function and tumorigenesis [17].

"So what." Non-coding RNAs, long and short, have gained more and more attention recently. They play an important role in gene regulation, and may count for the increased complexity in higher organisms including humans[18-20]. At present, very little is known about the function and mechanism of non-coding RNAs, and so it is of great scientific interest to study them. My goal in this study is to test whether certain noncoding RNAs have an effect on genome instability and DNA rearrangements. Abnormal transcripts seen in cancer were always considered a product of genome rearrangement in the past. But recently, RNA trans-splicing was found to be more prevalent in humans than previously thought[20-22], and some cancer-associated fusion RNAs are also detected in normal cells[21]. All these point to a possibility that chimeric RNA can be an initiator, rather than a by-product of genome rearrangement[23]. This possibly gains support from studies in *Oxytricha*, where programmed RNA-mediated genome rearrangement occurs during development[10]. By searching for "abnormal" RNA transcripts in both breast cancer cells and normal tissues, and by studying a class of small RNAs that functions during genome rearrangements, I hope to contribute to our understanding of genome instability, which may benefit breast cancer diagnosis and intervention.

References

1 Aguilera, A. & Gomez-Gonzalez, B. Genome instability: a mechanistic view of its causes and consequences. *Nat Rev Genet* **9**, 204-217 (2008).
2 Sieber, O. M., Heinimann, K. & Tomlinson, I. P. Genomic instability--the engine of tumorigenesis? *Nat Rev Cancer* **3**, 701-708 (2003).
3 Ingvarsson, S. Genomic instability and breast cancer progression. *Cancer genomics proteomics* **3**, 137 (2006).
4 Volik, S. End-sequence profiling: Sequence-based analysis of aberrant genomes. *Proceedings of the National Academy of Sciences of the United States of America* **100**, 7696 (2003).
5 Volik, S. Decoding the fine-scale structure of a breast cancer genome and transcriptome. *Genome research* **16**, 394 (2006).
6 Campbell, P. J. *et al.* Identification of somatically acquired rearrangements in cancer using genome-wide massively parallel paired-end sequencing. *Nat Genet* **40**, 722-729 (2008).
7 Hampton, O. A. *et al.* A sequence-level map of chromosomal breakpoints in the MCF-7 breast cancer cell line yields insights into the evolution of a cancer genome. *Genome Res* (2008).
8 Ruan, Y. *et al.* Fusion transcripts and transcribed retrotransposed loci discovered through comprehensive transcriptome analysis using Paired-End diTags (PETs). *Genome Res* **17**, 828-838 (2007).
9 Guffanti, A. *et al.* A transcriptional sketch of a primary human breast cancer by 454 deep sequencing. *BMC genomics* **10**, 163 (2009).
10 Nowacki, M. *et al.* RNA-mediated epigenetic programming of a genome-rearrangement pathway. *Nature* **451**, 153-158 (2008).
11 Edwards, S. L. *et al.* Resistance to therapy caused by intragenic deletion in BRCA2. *Nature* **451**, 1111-1115 (2008).
12 Storici, F., Bebenek, K., Kunkel, T. A., Gordenin, D. A. & Resnick, M. A. RNA-templated DNA repair. *Nature* **447**, 338-341 (2007).
13 Shen, Y. *et al.* RNA-driven genetic changes in bacteria and in human cells. *Mutat Res* **717**, 91-98, doi:10.1016/j.mrfmmm.2011.03.016 (2011).
14 Barlund, M. *et al.* Cloning of BCAS3 (17q23) and BCAS4 (20q13) genes that undergo amplification, overexpression, and fusion in breast cancer. *Genes Chromosomes Cancer* **35**, 311-317 (2002).
15 McPherson, A. *et al.* deFuse: an algorithm for gene fusion discovery in tumor RNA-Seq data. *PLoS Comput Biol* **7**, e1001138, doi:10.1371/journal.pcbi.1001138
10-PLCB-RA-2589R4 [pii] (2011).
16 Ye, Y. *et al.* Identification of Piwil2-like (PL2L) proteins that promote tumorigenesis. *PLoS One* **5**, e13406.
17 Liu, J. J. *et al.* Piwil2 is expressed in various stages of breast cancers and has the potential to be used as a novel biomarker. *Int J Clin Exp Pathol* **3**, 328-337.
18 Costa, F. F. Non-coding RNAs, epigenetics and complexity. *Gene* **410**, 9-17 (2008).
19 Kawaji, H. *et al.* Hidden layers of human small RNAs. *BMC genomics* **9**, 157 (2008).
20 Viles, K. D. & Sullenger, B. A. Proximity-dependent and proximity-independent trans-splicing in mammalian cells. *RNA (New York, N.Y* **14**, 1081-1094 (2008).
21 Li, H., Wang, J., Mor, G. & Sklar, J. A neoplastic gene fusion mimics trans-splicing of RNAs in normal human cells. *Science* **321**, 1357-1361 (2008).

22 Herai, R. H. & Yamagishi, M. E. Detection of human interchromosomal trans-splicing in sequence databanks. *Briefings in bioinformatics* **11**, 198-209.
23 Rowley, J. D. & Blumenthal, T. Medicine. The cart before the horse. *Science* **321**, 1302-1304 (2008).

Appendices

Detection of a common chimeric transcript between human chromosomes 7 and 16

Wenwen Fang[1], Yong Wei[1], Yibin Kang[1], and Laura F. Landweber[2]*

[1]Department of Molecular Biology, [2]Department of Ecology and Evolutionary Biology, Princeton University, NJ 08544, USA.

Email addresses:

wenwenf@princeton.edu;yongwei@princeton.edu; ykang@princeton.edu; lfl@princeton.edu.

*To Whom Correspondence Should Be Addressed.

Abstract
Interchromosomal chimeric RNA molecules are often transcription products from genomic rearrangement in cancerous cells. Here we report the computational detection of an interchromosomal RNA fusion between *ZC3HAV1L* and *CHMP1A* from RNA-seq data of normal human mammary epithelial cells, and experimental confirmation of the chimeric transcript in multiple human cells and tissues. Our experimental characterization also detected three variants of the *ZC3HAV1L-CHMP1A* chimeric RNA, suggesting that these genes are involved in complex splicing. The fusion sequence at the novel exon-exon boundary, and the absence of corresponding DNA rearrangement suggest that this chimeric RNA is likely produced by *trans*-splicing in human cells.

Keywords
Chimeric transcripts; RNA fusion; *trans*-splicing; genome rearrangement

Findings
High-throughput sequencing techniques have allowed characterization of genome and transcriptome catalogs in an unprecedented detail, revealing complex structures of genome rearrangements [1] and transcript networks of chimeric RNAs in human cells [2]. Because genome instability is a hallmark of cancer [3], most studies of genome rearrangements and RNA fusions focus on caner cells, and some chimeric RNAs appear to result from DNA rearrangements [4-8]. In fact, some studies use RNA-seq data as a guide to annotate genome rearrangement [9]. Several findings, however, suggest that normal human cells also produce chimeric RNA through *trans*-splicing [10-13]. Notably, Li *et al.* demonstrated that in normal endometrial cells, *trans*-splicing produces a chimeric RNA that is identical to a fusion transcript present in endometrial stromal tumor cells [13]. The corresponding chromosomal translocation which may permit production of this chimeric RNA by *cis*-splicing is present in endometrial tumor cells, but not detectable in normal endometrial cells. This hints at the possibility that RNA fusion may even predispose relevant genomic loci to rearrangement [13, 14], via RNA-guided DNA recombination, which our lab previously discovered in the ciliate *Oxytricha* [15].

We therefore asked whether we can identify more occurrences of chimeric transcripts, especially those involving genes on separate chromosomes, in normal human cells. We mined high-throughput RNA-seq data for human mammary epithelial cells (HMEC) available from the ENCODE project [16]. The deFuse program [17] predicted one interchromosomal RNA fusion between the genes encoding ZC3HAV1L (zinc finger CCCH-type, antiviral 1-like) and CHMP1A (charged multivesicular body protein 1A), located on chromosomes 7 and 16, respectively (Figure 1A). *ZC3HAV1L* contains 5 exons and encodes a 300 residue protein. *CHMP1A* has two protein-coding transcript isoforms, according to NCBI annotation; transcript variant 1 contains 6 exons encoding a 240 residue protein, and transcript variant 2 contains 7 exons encoding a 196 residue protein, which functions as a tumor suppressor in human kidney and pancreas [18, 19]. *CHMP1A* isoform 1 skips exon 2, but has a larger exon 7. Here we used *CHMP1A* isoform 1 as the reference for annotation purposes, because the fusion product we detect contains the larger exon 7 (see below).

Because we initially predicted the presence of the *ZC3HAV1L-CHMP1A* fusion in a breast cell line, we first verified the presence of this chimeric transcript in human mammary cells. Use of a primer pair that amplifies across the predicted *ZC3HAV1L-CHMP1A* fusion junction (from *ZC3HAV1L* exon 2 to *CHMP1A* exon 6) confirms the presence of this fusion at the RNA level in HMLE cells, which are human mammary cells derived from HMEC, but not at the DNA level from matching genomic DNA (Figure 1B). Sequencing of the PCR product verified the fusion junction (Figure 1C). The same PCR analysis suggests that the *ZC3HAV1L-CHMP1A* fusion is present in MCF10A, an immortalized but otherwise normal human mammary epithelial cell lines, as well as two human breast cancer cell lines, MB-MDA231 and MCF7. We also detected the fusion in CAPAN1, a pancreatic cancer cell line, human embryonic kidney (HEK) 293T cells, and CEMT, a human T cell line. In addition, we were able to amplify the fusion RNA from commercially-available human universal reference RNA and a panel of human tissue RNAs (Figure 1B). These results suggest that the *ZC3HAV1L-CHMP1A* chimeric RNA is common across multiple human tissue types, both healthy and diseased.

Curiously, we detected some minor PCR products of different sizes as well. Sequencing revealed that some of them reflect alternative splicing of the *ZC3HAV1L-CHMP1A* chimeric RNA, adding additional complexity to this fusion transcript. A common splicing variant present in multiple tissue types is the full-length chimeric RNA skipping *ZC3HAV1L* exon 3 (Figure 1B, D). Two other minor RT-PCR products suggest splicing between partial intron 2 and exon 3, and partial exon 3 and intron 3 of *ZC3HAV1L*, respectively (Figure 1B, D). Sequence alignments (see Additional file 1) indicate that these alternative splicing events all occur at canonical splicing sites, suggesting that they likely derive from authentic, alternative RNA splicing, rather than an *in vitro* RT-PCR artifact. A fusion product detected from RT-PCR analysis of human prostate RNA fuses *ZC3HAV1L* exon 4 to part of *CHMP1A* exon 6, not at canonical splicing site but between a pair of 5 bp direct repeats at the boundary. Therefore, this may represent either an endogenous activity or just template-switching during the reverse transcription step in our experimental procedure.

The major chimeric, fusion RNA that joins exon 4 of *ZC3HAV1L* to exon 5 of *CHMP1A* preserves the open reading frame. We therefore tested whether the entire open reading frame could be detected from mRNA by nested PCR. The use of primers located upstream and downstream of the predicted start and stop codons in the first round of PCR, and then a nested primer pair between the start and stop codons did amplify a product containing the predicted open reading frame, as well as the *ZC3HAV1L* exon3^{-} version of the transcript, though at much lower levels, consistent with our previous PCR result (Figure 1E). We infer that the major chimeric RNA that joins *ZC3HAV1L* and *CHMP1A* may encode a novel fusion protein.

From qPCR analysis, we estimated that *the ZC3HAV1L-CHMP1A* chimeric RNA is present at ~ 0.1 copies per HMLE cell, suggesting that its expression is limited to either a small population of cells, or a transient time window. We also assayed the relative abundance of this chimeric RNA compared to *beta-actin*, a constitutively expressed gene, and found that the relative levels of the *ZC3HAV1L-CHMP1A* chimeric RNA differ in different samples (Figure 1F). This suggests that the production of *ZC3HAV1L-CHMP1A* RNA might be regulated, or it might be a stochastic event.

In summary, we report the discovery of a chimeric RNA between *ZC3HAV1L* and *CHMP1A* in human, located on chromosome 7 and 16, respectively. The fusion occurs at an exon-exon boundary, and was detected both computationally and experimentally from different cells or tissue types. This suggests it is not an artifact from reverse transcription, and is likely an authentic *trans*-splicing product. We also detected three minor variants which also likely result from *trans*-splicing, because the fusion occurs at canonical "GT-AG" splicing sites. The fusion products are present at very low levels, and thus may reflect promiscuous splicing involving *ZC3HAV1L* and *CHMP1A*.

Could such low abundance chimeric RNAs have any function? While the major chimeric RNA that we detected preserves open reading frame and could potentially produce a novel fusion protein or proteins, we propose that such examples of chimeric RNA may also occasionally impact somatic genome rearrangements, facilitating rogue recombination events between the two respective chromosomes [14, 15]. The ability of RNA to influence genome remodeling has gained considerable support and interest over the past few years [20], but RNA-guided DNA rearrangement in humans still needs further investigation, emphasizing the importance of detecting more chimeric RNAs and their possible DNA rearrangements in normal or diseased tissue.

Methods

Computational detection of fusion RNAs

HMEC polyA-RNA-seq data from ENCODE project (http://hgdownload.cse.ucsc.edu/goldenPath/hg19/encodeDCC/wgEncodeCshlLongRnaSeq/) were downloaded from UCSC genome browser website. DeFuse 0.4.3 [17] was used to detect fusion transcripts, requiring the presence of two pairs of spanning reads and one split read at the junction. We divided the RNA-seq data to subsets of 10 and 20 million paired-end reads to accommodate computation memory, and the *ZC3HAV1L-CHMP1A* fusion was predicted in two independent subsets of data.

RNA and genomic DNA extraction

Total RNA from HMLE, MB-MDA231, MCF7, CAPAN1, MCF10A, HEK293T, and CEMT cells was extracted using RNeasy (Qiagen), and DNase treated with TURBO DNA-free kit (Ambion) following the manufacturer's instructions. Human reference RNA was purchased from Stratagene, and the human tissue RNA panel was purchased from Clontech. Genomic DNA was extracted using NucleoSpin Tissue (Macherey-Nagel).

cDNA synthesis

3.5 μg of RNA was reverse transcribed with SuperScriptIII reverse transcriptase (Invitrogen), in a 20 μl reaction following the oligo(dT) priming protocol.

PCR analysis

FastStart High Fidelity PCR system (Roche) was used to amplify fusion product from 1μl of cDNA (equivalent of 175 ng RNA), or 200 ng genomic DNA. To detect the predicted fusion from *ZC3HAV1L* exon 2 to *CHMP1A* exon 6, the following program was used: 95 °C 2 min initial denaturing; 95 °C 30 s, 58 °C 30 s, 72 °C 45 s for 36 cycles; 72 °C 7 min. To recover the

entire coding region by nested PCR, the following program was used. First round: 95 °C 2 min initial denaturing; 95 °C 30 s, 60 °C 30 s, 72 °C 90 s for 20 cycles; 72 °C 7 min. The PCR reaction was diluted 100 fold, and used in the second round of PCR: 95 °C 2 min initial denaturing; 64 °C 30 s, 72 °C 80 s for 30 cycles; 72 °C 7 min.

Sanger sequencing of PCR products
PCR products were either sent for direct Sanger sequencing (Genewiz) following Genewiz DNA sequencing instructions, or TOPO-cloned (Invitrogen) for colony PCR and sequencing (Genewiz).

qPCR analysis
The 7900HT Fast Real-Time PCR System and SYBR green master mix (Applied Biosystems) were used for qPCR analysis with the default cycling program. Standard curves for each primer pair were generated with five ten-fold serial dilutions of appropriate control plasmids in yeast RNA, allowing absolute quantification of DNA levels. Primer specificity was confirmed by a melt-curve analysis. Each primer pair detected the full range of standards with a correlation of $R^2 > 0.99$.

Oligonucleotide sequences (5'-3')

ZC3HAV1L_exon2_F	TGGTCTCAATGAAAACCAGCTTCGG
CHMP1A_exon6_R	ATTCTCCTCGGCGATCTGCATGATG
ZC3HAV1L_nested_F	AGCGACCATGGCGGAGCCCA
CHMP1A_nested_R	CACCGCCCAACCTAAAAGAACAGG
ZC3HAV1L_cds_F	ATGGCGGAGCCCACAGTGTGCTCC
CHMP1A_cds_R	CTAAGGCCACGCAGGCCTGGCAG
ZC3HAV1L_qPCR_F	AGAAGCTGGTCCTCTGGCTTCTGT
CHMP1A_qPCR_R	TCACCTGGGCCATATTCTTGGTCA
ACTB_qPCR_F	GCACAGAGCCTCGCCTT
ACTB_qPCR_R	CCTTGCACATGCCGGAG

List of abbreviations used
ZC3HAV1L: zinc finger CCCH-type, antiviral 1-like
CHMP1A: charged multivesicular body protein 1A
HMEC: human mammary epithelial cell
HEK293T: human embryonic kidney 293T
qPCR: quantitative PCR
gDNA: genomic DNA

Competing interests
The authors declare that they have no competing interests.

Authors' contributions
W.F. performed computational analysis of RNA-seq data, and experimental characterization of the chimeric RNAs. All authors participated in project design and manuscript writing.

Acknowledgements
We thank data producers from the ENCODE Consortium for HMEC RNA-seq data; R. Weinberg lab (Whitehead Institute) for sharing HMLE cells, and T. Li for sharing HEK293T and CEMT cells; E. Dolzhenko, A. D. Goldman and X. Chen for computational advice. This study was supported by NIH grant GM59708 and NSF grants 0923810 and 0900544 (to L.F.L.); Brewster Foundation, the Champalimaud Foundation, NIH grants R01CA134519 and R01CA141062 (to Y.K.); and a DOD BCRP pre-doctoral fellowship W81XWH-10-1-0122 (to W.F.).

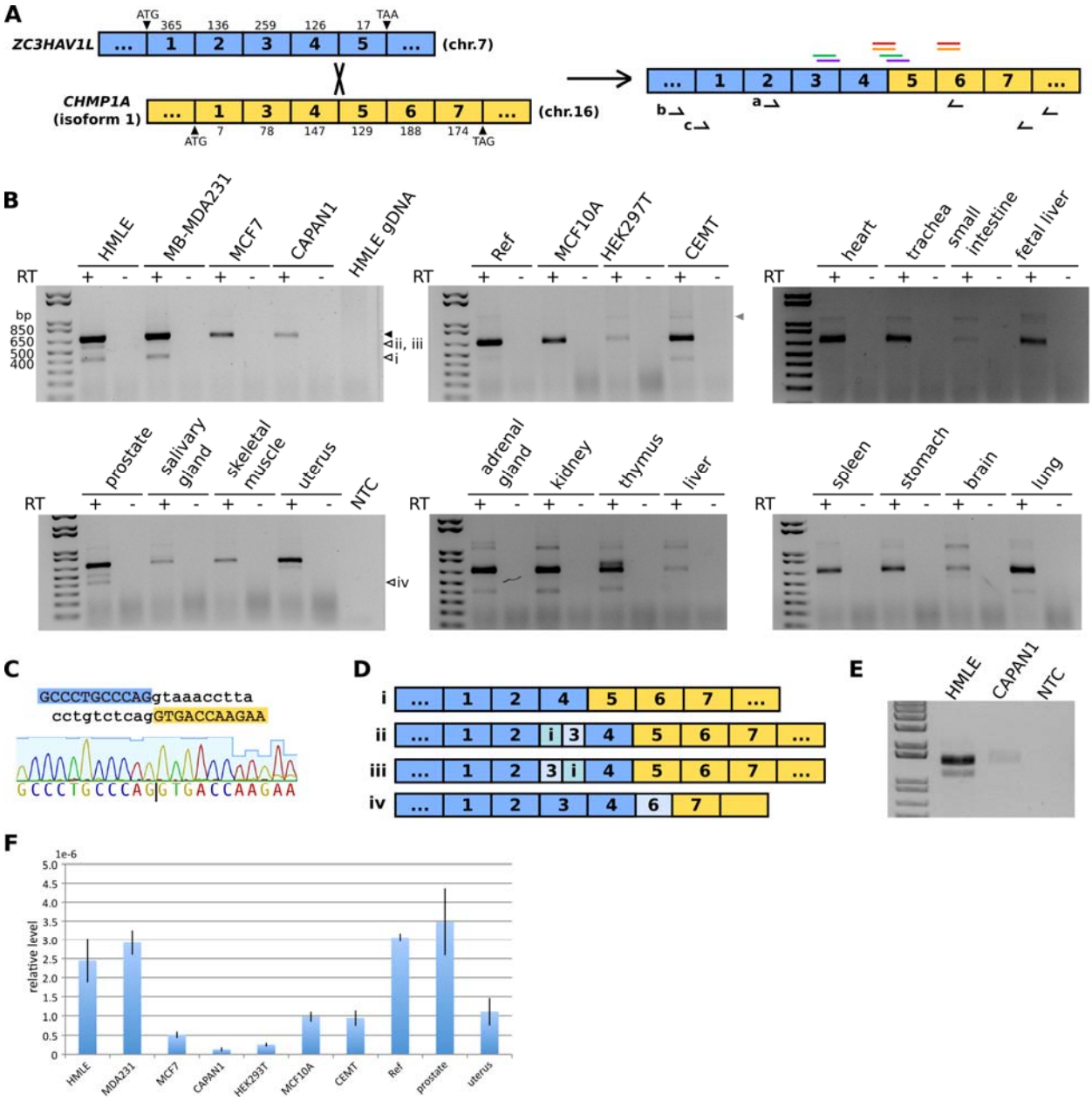
References
1. Maher CA, Wilson RK: **Chromothripsis and human disease: piecing together the shattering process.** *Cell* 2012, **148:**29-32.
2. Djebali S, Lagarde J, Kapranov P, Lacroix V, Borel C, Mudge JM, Howald C, Foissac S, Ucla C, Chrast J, et al: **Evidence for transcript networks composed of chimeric RNAs in human cells.** *PLoS One* 2012, **7:**e28213.
3. Hanahan D, Weinberg RA: **Hallmarks of cancer: the next generation.** *Cell* 2011, **144:**646-674.
4. Pflueger D, Terry S, Sboner A, Habegger L, Esgueva R, Lin PC, Svensson MA, Kitabayashi N, Moss BJ, MacDonald TY, et al: **Discovery of non-ETS gene fusions in human prostate cancer using next-generation RNA sequencing.** *Genome Res* 2011, **21:**56-67.
5. Maher CA, Kumar-Sinha C, Cao X, Kalyana-Sundaram S, Han B, Jing X, Sam L, Barrette T, Palanisamy N, Chinnaiyan AM: **Transcriptome sequencing to detect gene fusions in cancer.** *Nature* 2009, **458:**97-101.
6. Maher CA, Palanisamy N, Brenner JC, Cao X, Kalyana-Sundaram S, Luo S, Khrebtukova I, Barrette TR, Grasso C, Yu J, et al: **Chimeric transcript discovery by paired-end transcriptome sequencing.** *Proc Natl Acad Sci U S A* 2009, **106:**12353-12358.
7. Edgren H, Murumagi A, Kangaspeska S, Nicorici D, Hongisto V, Kleivi K, Rye IH, Nyberg S, Wolf M, Borresen-Dale AL, Kallioniemi O: **Identification of fusion genes in breast cancer by paired-end RNA-sequencing.** *Genome Biol*, **12:**R6.
8. Ruan Y, Ooi HS, Choo SW, Chiu KP, Zhao XD, Srinivasan KG, Yao F, Choo CY, Liu J, Ariyaratne P, et al: **Fusion transcripts and transcribed retrotransposed loci discovered through comprehensive transcriptome analysis using Paired-End diTags (PETs).** *Genome Res* 2007, **17:**828-838.
9. Zhao Q, Caballero OL, Levy S, Stevenson BJ, Iseli C, De Souza SJ, Galante PA, Busam D, Leversha MA, Chadalavada K, et al: **Transcriptome-guided characterization of genomic rearrangements in a breast cancer cell line.** *Proceedings of the National Academy of Sciences of the United States of America* 2009, **106:**1886-1891.

10. Li BL, Li XL, Duan ZJ, Lee O, Lin S, Ma ZM, Chang CC, Yang XY, Park JP, Mohandas TK, et al: **Human acyl-CoA:cholesterol acyltransferase-1 (ACAT-1) gene organization and evidence that the 4.3-kilobase ACAT-1 mRNA is produced from two different chromosomes.** *J Biol Chem* 1999, **274:**11060-11071.
11. Hahn Y, Bera TK, Gehlhaus K, Kirsch IR, Pastan IH, Lee B: **Finding fusion genes resulting from chromosome rearrangement by analyzing the expressed sequence databases.** *Proc Natl Acad Sci U S A* 2004, **101:**13257-13261.
12. Herai RH, Yamagishi ME: **Detection of human interchromosomal trans-splicing in sequence databanks.** *Brief Bioinform* 2010, **11:**198-209.
13. Li H, Wang J, Mor G, Sklar J: **A neoplastic gene fusion mimics trans-splicing of RNAs in normal human cells.** *Science* 2008, **321:**1357-1361.
14. Rowley JD, Blumenthal T: **Medicine. The cart before the horse.** *Science* 2008, **321:**1302-1304.
15. Nowacki M, Vijayan V, Zhou Y, Schotanus K, Doak TG, Landweber LF: **RNA-mediated epigenetic programming of a genome-rearrangement pathway.** *Nature* 2008, **451:**153-158.
16. Birney E, Stamatoyannopoulos JA, Dutta A, Guigo R, Gingeras TR, Margulies EH, Weng Z, Snyder M, Dermitzakis ET, Thurman RE, et al: **Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project.** *Nature* 2007, **447:**799-816.
17. McPherson A, Hormozdiari F, Zayed A, Giuliany R, Ha G, Sun MG, Griffith M, Heravi Moussavi A, Senz J, Melnyk N, et al: **deFuse: an algorithm for gene fusion discovery in tumor RNA-Seq data.** *PLoS Comput Biol* 2011, **7:**e1001138.
18. You Z, Xin Y, Liu Y, Sun J, Zhou G, Gao H, Xu P, Chen Y, Chen G, Zhang L, et al: **Chmp1A acts as a tumor suppressor gene that inhibits proliferation of renal cell carcinoma.** *Cancer Lett* 2012, **319:**190-196.
19. Li J, Belogortseva N, Porter D, Park M: **Chmp1A functions as a novel tumor suppressor gene in human embryonic kidney and ductal pancreatic tumor cells.** *Cell Cycle* 2008, **7:**2886-2893.
20. Nowacki M, Shetty K, Landweber LF: **RNA-Mediated Epigenetic Programming of Genome Rearrangements.** *Annual review of genomics and human genetics* 2011, **12:**367-389.

Figure legends

Figure 1. Detection of *ZC3HAV1L-CHMP1A* chimeric RNA in human cells. (A) Schematic representation of the *ZC3HAV1L-CHMP1A* chimeric RNA. Blue and yellow boxes indicate exons from *ZC3HAV1L* and *CHMP1A*, respectively. Above the predicted fusion, colored bars indicate paired reads from the ENCODE HMEC RNA-seq data. The same color indicates a read pair supporting the fusion. Arrows below the predicted fusion indicate primer pairs used in PCR analyses (primer pair *a* for PCR in panel B, and primer pairs *b* and *c* for nested PCR in panel E). (B) RT-PCR detection of the *ZC3HAV1L-CHMP1A* chimeric RNA in multiple human cells and tissues. The filled arrowhead indicates the major predicted fusion product; open arrowheads indicate minor alternative chimeric transcripts (i-iv). "RT -" indicates no reverse transcriptase negative controls. "NTC" indicates no template negative control for PCR analysis. (C) Junction sequence of the *ZC3HAV1L-CHMP1A* chimeric RNA. Blue and yellow shading highlights sequences from the end of *ZC3HAV1L* exon 4 and start of *CHMP1A* exon 5, respectively. Lower case letters indicate intron sequences. The chromatogram shows Sanger sequencing results at the junction from RT-PCR analysis of HMLE cells. (D) Schematic representation of minor alternative *ZC3HAV1L-CHMP1A* chimeric RNAs detected from PCR analyses shown in (B). Partial exon and introns ("i") are indicated by different shades of color. Sequence alignment for the major and minor chimeric RNAs is provided in an additional file. (E) *ZC3HAV1L-CHMP1A* fusion may encode a novel protein. Shown is nested RT-PCR amplification of the predicted fusion coding sequence from HMLE and CAPAN1 cells. The larger band indicates the expected PCR product, whereas the smaller band in HMLE lane indicates the *ZC3HAV1L* exon3^{-} variant, which does not preserve open reading frame. (F) *ZC3HAV1L-CHMP1A* chimeric RNA levels differ in different human cell and tissue types. The copy number of the RNA fusion between *ZC3HAV1L* exon 4 and *CHMP1A* exon 5 is normalized against *beta-actin* mRNA.

Figure 1


A
ZC3HAV1L
CHMP1A (isoform 1)
(chr.7)
(chr.16)
ATG
TAA
B
HMLE
MB-MDA231
MCF7
CAPAN1
HMLE gDNA
Ref
MCF10A
HEK297T
CEMT
heart
trachea
small intestine
fetal liver
prostate
salivary gland
skeletal muscle
uterus
NTC
adrenal gland
kidney
thymus
liver
spleen
stomach
brain
lung
RT
bp
850
650
500
400
i
ii, iii
iv
C
GCCCTGCCCAGgtaaacctta
cctgtctcagGTGACCAAGAA
D
E
HMLE
CAPAN1
NTC
F
relative level
HMLE
MDA231
MCF7
CAPAN1
HEK293T
MCF10A
CEMT
Ref
prostate
uterus

Piwi-Interacting RNAs Protect DNA Against Loss During *Oxytricha* Genome Rearrangement

Wenwen Fang[1], Xing Wang[2], John R. Bracht[2], Mariusz Nowacki[3], Laura F. Landweber[2]*

[1]Department of Molecular Biology, [2]Department of Ecology and Evolutionary Biology, Princeton University, NJ 08544, USA.

[3]Institute of Cell Biology, University of Bern, 3012 Bern, Switzerland

*To whom correspondence should be addressed.

Tel. 609-258-1947

Fax 609-258-7892

Email: lfl@princeton.edu

Summary

Genome duality in ciliated protozoa offers a unique system to showcase their epigenome as a model of inheritance. In *Oxytricha,* the somatic genome is responsible for vegetative growth, while the germline contributes DNA to the next sexual generation. Somatic nuclear development removes all transposons and other so-called "junk DNA", which comprise ~95% of the germline. We demonstrate that Piwi-interacting small RNAs (piRNAs) from the maternal nucleus can specify genomic regions for retention in this process. *Oxytricha* piRNAs map primarily to the somatic genome, representing the ~5% of the germline that is retained. Furthermore, injection of synthetic piRNAs corresponding to normally-deleted regions leads to their retention in later generations. Our findings highlight small RNAs (sRNAs) as powerful transgenerational carriers of epigenetic information for genome programming.

Introduction

Oxytricha trifallax harbors two types of nuclei within its single cell: a somatic macronucleus and a germline micronucleus. In each sexual conjugation cycle, the old somatic nucleus disintegrates and a new soma develops from the germline. The germline genome contains large quantities of so-called “junk”, including transposons, satellite repeats, intergenic sequences, and internally eliminated sequences [IESs], all of which undergo programmed deletion during somatic nuclear development. This compresses the ~1Gb germline genome to a gene-dense somatic genome of only ~50Mb. Severe chromosome fragmentation necessitates events that fuse and reorder the tens of thousands of gene pieces remaining [macronuclear-destined sequences, MDSs]. Lastly, telomere addition and amplification produce mature macronuclear “nanochromosomes”, typically ~ 3 kb bearing just one gene (Prescott, 2000; Swart et al., in revision).

The precise and reproducible DNA rearrangements in *Oxytricha*, as well as other ciliates, show a prominent pattern of maternal inheritance (Chalker and Yao, 1996; Duharcourt et al., 1995) mediated by non-coding RNAs (Lepere et al., 2008; Mochizuki et al., 2002; Nowacki et al., 2008; Yao et al., 2003). In the distantly related ciliates *Tetrahymena* and *Paramecium*, a class of small RNAs (sRNAs) called scanRNAs (scnRNAs) derive from the germline and scan the macronuclear genome or transcriptome for sequence identity, leaving sRNAs that lack matches in the macronucleus (and thus are micronuclear-limited) to target germline-specific DNA for deletion during nuclear development (Lepere et al., 2008; Mochizuki et al., 2002). We previously demonstrated that long non-coding RNA templates from the maternal somatic nucleus guide whole-genome reorganization in *Oxytricha* (Nowacki et al., 2008). However, abundant ~27 nucleotide (nt) sRNAs also accumulate during conjugation (Figure 1A, B), suggesting a role for sRNA pathways during *Oxytricha* genome rearrangement, which we investigate here.

Small RNAs, in association with an Argonaute/Piwi family protein, can act as sequence-specific guides to regulate gene expression and chromatin structure (Bartel, 2004; Zamore and Haley, 2005). The subclass of Piwi-interacting RNAs (piRNAs) is abundant in the animal germline, and its best-understood function is transposon silencing (Brennecke et al., 2007; Carmell et al., 2007; Grimson et al., 2008; Houwing et al., 2007; Malone and Hannon, 2009; Saito et al., 2006; Vagin et al., 2006). In *Drosophila*, maternally-deposited piRNAs can prime piRNA production and transposon immunity in daughters, underscoring a role for piRNAs as epigenetic information carriers (Brennecke et al., 2008). Much less is known about piRNAs that do not map to transposon or repetitive DNA, despite the fact that such piRNAs can be

overrepresented in mammalian testis, for example (Aravin et al., 2006; Girard et al., 2006; Grivna et al., 2006; Lau et al., 2006; Watanabe et al., 2006). Deep sequencing revealed piRNAs that map to both 3' UTR and coding regions of unique sets of transcripts, suggesting a role in gene expression regulation (Gan et al., 2011; Robine et al., 2009). However, functional evidence is lacking, except for case studies (Rajasethupathy et al., 2012; Rouget et al., 2010; Saito et al., 2009). A recent study in plant and human cells uncovered a surprising role for sRNAs in double-stranded DNA break repair, relating sRNA function to genome integrity (Wei et al., 2012). Here, by studying *Oxytricha*, a microbial eukaryote that serves as a paragon for investigations of genome rearrangement, we suggest a genome-wide protective role for piRNAs that map to genic regions. We show that *Oxytricha* piRNAs derive from the somatic genome, and localization of the associated Piwi protein shifts from the maternal to the developing somatic nucleus during genome rearrangement. Furthermore, injection of sRNAs targeting normally-deleted regions leads to their retention in the sexual offspring across multiple generations, demonstrating a role for these piRNAs in transgenerational epigenetic inheritance.

Results

27 nt sRNAs and Otiwi1 co-express and interact during early conjugation.

A time-course examination of total cellular RNA identifies an abundant class of ~27 nt sRNAs exclusively expressed during early conjugation (Figure 1A). A much less abundant class of ~21 nt sRNAs are also detectable by radioactive labeling (Figure 1B), but present in both vegetative and conjugating cells (Figure 1B) and could be siRNAs involved in gene silencing, like the 23-24 nt siRNAs in *Tetrahymena* (Lee and Collins, 2006) and *Paramecium* (Lepere et al., 2009). Here, we focus on the development-specific 27 nt sRNAs. To probe the function of these sRNAs, we first analyzed Argonaute proteins in *Oxytricha*. Using *Tetrahymena* Twi1p protein as a query, we retrieved thirteen *Oxytricha* homologs (Otiwi1-13) from the macronuclear genome assembly (Swart et al., in revision). This number of Argonaute proteins is comparable to that in *Tetrahymena* (Couvillion et al., 2009) and *Paramecium* (Bouhouche et al., 2011), but higher than most metazoa, reflecting the richness of sRNA-based regulation in single-celled ciliates. The thirteen *Oxytricha* homologs form two clades (Figure 1C). Clade I (Otiwi1-4 and Otiwi11) groups more closely with Piwi subfamily proteins from metazoa and *Dictyostelium*, and their mRNA show elevated expression levels during conjugation, based on RT-PCR analysis (Figure 1D). In addition, expression levels of *Otiwi1* and *Otiwi4* are higher in early conjugation, while *Otiwi2*, *3*, and *11* are more restricted to late conjugation (Figure 1D). Expression of the Clade II *Otiwi* mRNAs, by contrast, is relatively constitutive across vegetative growth and conjugation (Figure 1D).

Use of an antibody that based on mass spectrometry primarily recognizes Otiwi1 (Figure S1, Table S1) confirmed that the Otiwi1 protein interacts with the ~27 nt sRNAs throughout early conjugation (Figure 1E). Based on this interaction, we denote Otiwi1-interacting sRNAs as a class of piRNAs in *Oxytricha*. Northern analysis suggests that *Otiwi1* mRNA abundance peaks 12 hr after mixing of mating types and then decreases as conjugation proceeds (Figure 1F). In western analysis, Otiwi1 protein levels become detectable at 12 hr, and then peak at 18-24 hr post mixing (Figure 1G), which fits well with the 27 nt piRNA expression pattern (Figure 1A, B).

Oxytricha piRNAs contain 5' monophosphate but no 3' end modification.

The ends of sRNAs differ among classes. Most microRNAs, siRNAs and piRNAs contain a 5' monophosphate, but secondary siRNAs in *C. elegans* (Pak and Fire, 2007; Sijen et al., 2007) and 27 nt sRNAs in *Entamoeba histolytica* (Zhang et al., 2008) are 5' polyphosphorylated. At the 3' end, plant microRNAs (Li et al., 2005), animal siRNAs (Ameres et al., 2010) and piRNAs (Houwing et al., 2007; Kirino and Mourelatos, 2007; Ohara et al., 2007; Saito et al., 2007) can be 2'-O-methylated, but animal microRNAs are not.

We therefore characterized end modifications of the Otiwi1-bound piRNAs in *Oxytricha*. Terminator exonuclease is specific to RNAs with a 5' monophosphate. *Oxytricha* piRNAs are sensitive to Terminator nuclease digestion, suggesting that they contain a 5′ monophosphate (Figure 1H). β-elimination post periodate treatment shortens RNAs without 3' end modifications by one nucleotide, whereas RNAs with 3' end modifications are resistant to β-elimination and thus remain the same size. *Oxytricha* piRNAs are not resistant to periodate oxidation, indicating a lack of 3′ end modifications (Figure 1I). This is surprising because the stabilizing 3' end methylation is conserved in animal piRNAs (Houwing et al., 2007; Kirino and Mourelatos, 2007; Ohara et al., 2007; Saito et al., 2007) as well as *Tetrahymena* scnRNAs (Kurth and Mochizuki, 2009). It is possible that *Oxytricha* piRNAs represent either a more ancestral state prior to the addition of 3' end modification, or that the end modification was lost in this lineage.

Structural studies of human Piwi PAZ domains suggest that a Piwi-specific insertion, not present in human Ago1, may facilitate formation of a larger binding pocket for the 3'-end of RNA, permitting the Piwi PAZ domain to accommodate the 3'-end methyl group on a piRNA (Tian et al., 2011). An alignment of Otiwi1 with human Piwi and Ago1 proteins suggests that Otiwi1 lacks this insertion (Figure 1J), consistent with the lack of 3'-methylation on *Oxytricha* 27 nt piRNAs.

Otiwi1 localization shifts from parental to developing macronucleus.

If piRNAs are important for *Oxytricha* genome rearrangement, we would expect to find the Otiwi1-piRNA complex in the developing macronucleus. Immunofluorescence suggests that Otiwi1 is absent from early conjugating pairs, when micronuclei begin meiosis (Figure 2A, see also Figure S2A for details of the *Oxytricha* sexual cycle), but Otiwi1 protein starts to accumulate in the maternal macronucleus in mid-to-late pairing stage (Figure 2B). Then Otiwi1 transiently localizes in both the cytoplasm and the developing macronucleus after fertilization (Figure 2C). Otiwi1 is abundant in the new somatic nucleus after pair separation (Figure 2D) and then its abundance decreases during late development (Figure 2E). This dynamic localization suggests that the Otiwi1-piRNA complex might transport piRNAs from the maternal nucleus to the developing nucleus where genome rearrangement occurs. In addition, a peptide competition assay using Otiwi1 and Otiwi4 C-terminal peptides further confirms the antibody's specificity for Otiwi1 (Figure S2B).

Otiwi1 is essential for the accumulation of 27 nt piRNAs and the viability of sexual progeny.

To test if *Otiwi1* is essential for new macronuclear development, we injected antisense phosphorothioate-backbone DNA oligonucleotides into cells 2-4 hr post-mixing of mating types. This significantly reduced *Otiwi1* mRNA and protein levels compared to control-oligonucleotide injections (Figure S3). The loss of protein signal after *Otiwi1* knockdown also confirms the

antibody's specificity. Notably, knockdown of *Otiwi1*, but not *Otiwi4*, results in depletion of the 27 nt piRNAs at 18 hr (Figure 3A), suggesting that Otiwi1 is the main interacting partner of this class of *Oxytricha* sRNAs. It is possible that piRNAs are not stable in the absence of *Otiwi1*, or alternatively, that the 27 nt piRNAs require *Otiwi1* for their biogenesis.

Otiwi1 knockdown cells appear normal at 12 hr post-mixing, but at 24 hr, 21% were arrested at pair stage, while only 1.8% were still pairs in control populations (Figure 3B-D). At 24 hr, those that appeared to be morphologically normal exconjugants (donut-shaped) were less active compared to control cells, and died between 24-48 hr post-mixing, with a dramatic decrease in survival between 24-28 hr (Figure 3B). We also observed significantly more ($p \leq 0.001$) cells with rounded and less transparent morphology (Figure 3G-J). These cells also show degeneration of the developing macronucleus, the old macronucleus and the micronucleus (Figure 3G-J), and eventually disintegrate. It is possible that these spherical structures might be an intermediate stage where cells degrade their nuclei before cell death. We conclude that *Otiwi1* is essential for exconjugant survival and new macronuclear development.

Oxytricha piRNAs map primarily to the somatic genome.

To identify the genomic source of piRNAs in *Oxytricha*, we deep sequenced sRNAs captured by Otiwi1 co-immunoprecipitation at 12, 19, 23, and 30 hr after mixing of mating types, early conjugation stages when both *Otiwi1* and piRNAs are expressed. Otiwi1-bound piRNAs are predominantly 27 nt (Figure 4A) and have a strong 5′-uridine (U) bias (Figure 4B). Sequencing of total sRNA at ~20 hr post-mixing of mating types suggests that the 27 nt 5'-U piRNAs are the major class of sRNAs during conjugation (Figure 4A, B), consistent with the RNA characterization in Figures 1A and B. Total sRNA also displays a minor peak of 21-22 nt (Figure 4A), which reflects the less abundant sRNA class detected by radioactive labeling (Figure 1B). Recently, Zahler et al. (2012) also deep sequenced total sRNA from vegetative and conjugating *Oxytricha* cells, which provides additional datasets for analysis.

Across all four time points, over 80% of Otiwi1-associated 27 nt 5'-U reads map to the *Oxytricha* macronuclear genome (Figure 4C), which represents ~5% of the micronuclear genome. Germline-limited TBE1 (Telomere-Bearing Elements) transposons occupy ~1% of the germline genome (Witherspoon et al., 1997), but only constitute less than 0.01% of the reads, making them considerably under-represented. Figure 4D shows representative piRNA mappings to typical macronuclear and micronuclear contigs, illustrating that piRNAs map only to MDS regions (see also Figure S4). Importantly, some piRNAs span MDS-MDS junctions (reads spanning cyan boxes in Figure 4D and Figure S4), suggesting that they do not derive from the micronucleus. To further exclude the possibility that these piRNAs might derive from distinct micronuclear piRNA loci, we mapped these 234,585 MDS-MDS junction-spanning sRNAs to a draft micronuclear genome assembly, sequenced to 50-fold coverage (Chen et al., unpublished data). Only 2 junction spanning piRNAs have a perfect match in the micronuclear genome of just 27 nt, which is not very different from 0.14 expected by chance, based on assumptions of sequence uniformity. This study therefore revealed no evidence of piRNA-specific loci in the micronuclear genome.

We do not observe any position bias in piRNA mapping, and a meta-gene analysis based on 10,497 fully-sequenced single-gene chromosomes (Swart et al., in revision) produces a nearly

uniform piRNA distribution across entire chromosomes, except for a small depletion toward the ends (Figure 4E, Figure S5A, D, G, J). At current sequencing depth, the 27 nt 5'-U reads cover 85-91% of these chromosomes at each of the four time points (Figure 4F, Figure S5B, E, H) and the combined reads from all four time points cover 96% (Figure S5K) with a near-equal representation of sense and antisense piRNAs (Figure 4G, Figure S5C, F, I, L). We conclude that the entire macronuclear genome may produce piRNAs. *Oxytricha* piRNAs either originate from the somatic nucleus or undergo selective amplification based on sequence matching to the macronuclear genome, resulting in their over-representation of sequences derived from the somatic genome.

Although piRNAs seem to map preferentially to coding regions compared to subtelomeric regions (Figure 4D) we do detect many piRNAs mapping to non-coding regions, such as 5' and 3' UTR and introns. Furthermore, the piRNA mapping pattern to *Otiwi3* at 12 hr post mixing, when there is no *Otiwi3* expression (Figure 1D), is indistinguishable from two constitutively expressed genes, *TEBPβ* and *actin I*. This argues against a correlation with mRNA expression for these piRNAs.

To further exclude correlation of piRNAs with mRNA levels, we analyzed piRNA mapping to 54 genes that show no expression in *Oxytricha* at any surveyed time point, based on RNA-seq data from vegetative cells, as well as 0, 10, 20, 40, and 60 hr post mixing of mating types. The mapping of 27 nt 5'-U piRNAs (all four piRNA datasets combined) shows no position bias relative to chromosome location, and the piRNA mapping "coverage" is just slightly lower than that from all 10,497 genes (93% vs. 96%, Figure S5N, O), which again suggests that the 27 nt piRNAs do not primarily derive from mRNA.

Telomere-containing sRNAs were present in the deep-sequenced library (~0.02%, Table S2) but excluded from mapping analysis. A separate analysis of sRNAs containing either "$G_4T_4G_4$" or "$C_4A_4C_4$" telomeric repeats suggests that "GT" telomeric reads may be piRNAs, because they are 27 nt with a 5'-U bias (data not shown), but "CA" telomeric reads are not. Given that telomere ends are typically 0.6% of a nanochromosome, the portion of telomeric reads is lower than expected if piRNAs are processed from RNA templates that are telomere-containing transcripts of macronuclear chromosomes (Nowacki et al., 2008). Because "GT" telomeric piRNAs usually start with non-repetitive sequence and end in G_4T_4 repeats, they more likely derive from macronuclear run-through pre-mRNA or mRNA transcripts, rather than TBE transposons, whose telomeric repeats are internal, or micronuclear telomeres with hundreds of such repeats (Prescott, 2000).

Injection of synthetic RNAs corresponding to normally deleted regions leads to their retention.

The data so far argue against a primary role for *Oxytricha* piRNAs in transposon silencing or DNA elimination. Instead, the predominant mapping of piRNAs to the macronuclear genome hints at an orthogonally-different pathway, in which *Oxytricha* piRNAs specify somatic regions for retention during DNA rearrangement. To directly test this possibility, we asked whether synthetic piRNAs whose sequences correspond to normally-deleted sequences (IESs) can guide their retention.

Synthetic piRNA design focused on short IESs that could be covered by a single piRNA. We also required a U at the first position, and chose regions of modestly higher GC content (25-30% GC vs. IES average 16.2%) to facilitate annealing. We injected synthetic 27 nt sRNA targeting an IES into 20-40 mating pairs at 12-18 hr post-mixing, pooled injected pairs and allowed them to finish conjugation and resume vegetative growth for ~1 week (5-8 divisions) before PCR analysis. In injection experiments targeting four different IESs in three genes (Contig22226.0 IES1, *actin I* IES4 and IES5, and *TEBPα* IES5, colored red in Figure 5A, B, C, D respectively; see also Figure S6A) the sRNA-injected cells specifically retained the IES whose sequence was present in the injected sRNA, and injection of either sense or antisense RNA protected the corresponding IES against deletion (Figure 5A-D, Figure S6A). Single-cell PCR analysis of Contig22226.0 after injection of a sRNA targeting IES1 further revealed that the majority of macronuclear chromosomes in 2 out of 3 examined cells displayed efficient IES retention (Figure S6B).

Injection of a degenerate 27 nt A/U–rich RNA containing all the features of *Oxytricha* piRNAs showed no IES retention. Furthermore, control injections with a 27 nt single-stranded DNA of the same sequence showed no IES retention in the progeny (Figure 5A-B) confirming that IES retention is a direct effect specific to RNA, and that the protective effect is not simply the consequence of hybridization and subtraction against other functional RNAs that guide DNA rearrangement.

The processing of other IESs in the same gene was unaffected by sRNA injection (Figure 5A-D and supplemental sequence alignment), suggesting that the protective effect requires sequence similarity. On the other hand, sRNAs containing two IES mismatches can have a protective effect, since an injected sRNA corresponding to a single allele also acted on different alleles of the same gene (Figure S6C). Deliberately-introduced substitutions on synthetic sRNAs never transferred to the retained DNA (Figure S6D, S6E), suggesting that the piRNAs are not the templates for RNA-guided DNA repair (Storici, 2008) during genome rearrangement, because substitutions can transfer from the long template RNAs (Nowacki et al., 2008). Instead we propose that the *Oxytricha* piRNAs prevent excision of a DNA segment.

The IES retention induced by sRNA injection is stable for at least 5 generations of asexual growth in all four injection experiments. We were also able to obtain clonal cell lines with Contig22226.0 IES1 retained, and these were stable for at least 20 doubling times and through an encycstment-excystment cycle (data not shown). Remarkably, when the offspring of injected cells either self-conjugate to produce F2 and F3 cells (Figure 5A) or are backcrossed to a wild-type parental strain (Figure 5B), the sexual offspring in the next generation also retain the IES. We therefore conclude that transient availability of *Oxytricha* piRNAs can effect heritable DNA sequence change in the somatic genome over multiple sexual generations.

To exclude the possibility that the injected RNA was still present in later generations, and to test whether IES-containing piRNAs were produced *de novo* in the next generation to mediate the IES retention effect, we deep sequenced sRNA at 19 hr post-mixing of the Contig22226.0 IES1^{+} strain and its wild-type parental strain. Out of 7 million reads, we detected two 5'-U IES-containing piRNAs (26 and 27 nt, respectively; Figure 5E, Figure S6F) neither of which was present in any wild-type sRNA pool, even though those were sequenced to much higher depth (161 million reads, combined). Furthermore, both RNA sequences are different from the original

injected synthetic RNA (Figure 5E), which is absent from the sequence pool, suggesting that the transgenerational effect of transient sRNA injection propagates via the production of new, endogenous IES-containing piRNAs. Our discovery that *Oxytricha* piRNAs can transfer epigenetic information across multiple sexual generations underscores the power of sRNA to influence genome programming and re-programming.

Discussion

piRNA-mediated protection against DNA loss

The most parsimonious model from our experiments is that *Oxytricha* piRNAs derive from the parental macronucleus, form complexes with Otiwi1 proteins, and inform the developing macronuclear genome which DNA sequences to retain (Figure 6). While this view is conceptually orthogonal to the model where piRNAs target transposon or non-self DNA for silencing or elimination, *Oxytricha* sRNAs nonetheless provide information to distinguish "self" versus "non-self", possibly a common theme in piRNA function. Intriguingly, two groups recently proposed that a class of 22 nt 5'-G sRNAs in *C. elegans* may act as an anti-silencing or "self" signal (Ashe et al., 2012; Lee et al., 2012; Shirayama et al., 2012). In *Ascaris*, a nematode that undergoes chromatin diminution, sRNAs do not target the eliminated repetitive sequences (Wang et al., 2011). It would be very informative to test if a similar protective role applies to *Ascaris* and other organisms, either with or without genome rearrangements.

Because piRNAs seem to accumulate relatively early in macronuclear development, and evidence suggests that DNA rearrangements occur later (Mollenbeck et al., 2008) it is possible that the ability to distinguish MDS versus IES information transfers from the piRNAs to chromatin at an early stage of macronuclear development, possibly via chromatin or base modifications. This type of epigenetic information transfer occurs in a wide range of organisms and biological contexts, including sRNA-mediated heteochromatin formation in *S. pombe* (Verdel et al., 2004; Volpe et al., 2002) and sRNA-guided DNA methylation in plants (Matzke et al., 2009; Wassenegger et al., 1994) and animals (Gu et al., 2012; Watanabe et al., 2011). A mechanism involving histone modification has been suggested downstream of scnRNA-guided DNA elimination in *Tetrahymena* (Liu et al., 2004; Liu et al., 2007). However, the length of a single MDS or IES in *Oxytricha* can be shorter than a nucleosome unit, which precludes the ability of a histone-based mechanism to direct precise DNA rearrangements in *Oxytricha*.

One mechanism we propose for the protective role of piRNAs against DNA deletion in *Oxytricha* is that piRNAs may "mark" MDSs and prevent cleavage at regions of sequence identity. If *Oxytricha* piRNAs help introduce DNA modifications, then these might serve, for example, as a signal to prevent TBE transposases that participate in DNA rearrangement (Nowacki et al., 2009) from introducing DNA breaks (Figure 6). This model suggests a new way in which piRNAs may antagonize transposon activity.

Roles for piRNAs and template RNAs in *Oxytricha* genome rearrangements

We previously reported that long RNA templates can both program MDS order and transfer point substitutions near MDS-MDS junctions from RNA to DNA (Nowacki et al., 2008). Here we propose that piRNAs specify retained regions (MDSs) in early genome rearrangement. Although long template RNA could be processed into sRNA, we propose that the two classes of noncoding RNA act independently for several reasons. First, while we can successfully reprogram the order of DNA segments with long RNA templates, we have been unable to do so with 27 nt sRNAs that span DNA recombination junctions, despite several attempts (data not shown). Second, artificial long RNA templates can transfer substitutions to the reprogrammed DNA sequence near junctions, presumably via RNA-guided DNA repair (Nowacki et al., 2008; Storici, 2008) but substitutions on injected piRNAs did not transfer (Figure S6D and S6E), consistent with a model where *Oxytricha* piRNAs confer DNA protection from rearrangement. These observations suggest that piRNAs are probably not templates for DNA repair and synthesis. In addition, we also detected fewer than expected telomere-containing piRNAs if telomere-to-telomere transcripts (Nowacki et al., 2008) are the main piRNAs precursors; however, we cannot exclude Illumina sequence bias producing a paucity of telomeric repeats. We suggest that piRNAs and the long template RNAs have distinct features and roles to ensure the integrity of the new somatic genome.

Evolution and comparison to other ciliate piRNAs

Ciliates are over one billion years old, with the genetic distance between *Oxytricha* and *Tetrahymena* close to that between humans and fungi (Parfrey et al., 2011). In terms of the Piwi-piRNA system, the conjugation-specific Piwi proteins show a similar expression and localization profile among *Oxytricha*, *Tetrahymena* and *Paramecium* (Bouhouche et al., 2011; Mochizuki et al., 2002), yet the source and function of *Oxytricha* piRNAs appear orthogonally different from that of *Tetrahymena* and *Paramecium*. ScnRNAs in *Tetrahymena* and *Paramecium* derive from the germline genome and ultimately target DNA for elimination (Lepere et al., 2008; Lepere et al., 2009; Mochizuki et al., 2002; Mochizuki and Gorovsky, 2004; Schoeberl et al., 2012); whereas *Oxytricha* piRNAs derive from the somatic genome and target sequences for retention. The opposite piRNA targeting in these species makes economic sense in both lineages, because in each case sRNAs predominantly target the minority class of the germline genome, with roughly 5% of the *Oxytricha* germline marked for preservation, as opposed to ~ 33% of the *Tetrahymena* germline that has to be marked for deletion (*Tetrahymena* Comparative Sequencing Project, Broad Institute of Harvard and MIT (http://www.broadinstitute.org/)). This evolutionary "sign change" suggests that piRNAs as sequence-dependent guides and their related pathways are extremely plastic on a deep evolutionary time scale, permitting the acquisition of new roles in diverse lineages to improve the efficiency of sRNA pathways or to alter the pathways themselves.

In addition to the drastic difference in biogenesis and function, there are two other major differences between *Oxytricha* piRNAs and *Tetrahymena* and *Paramecium* scnRNAs. First, *Paramecium* 25 nt scnRNAs have a 5'-UNG signature, and the non-5'-U scnRNAs have a "CNA" motif two nucleotides from the 3' end, indicating that scnRNAs form duplexes with 2 nt 3' overhangs, processed from dsRNA precursors by Dicer proteins (Lepere et al., 2009). This complementary signature was absent from *Oxytricha*, and it is unclear whether *Oxytricha* piRNAs derive from single- or double-stranded RNAs. In addition, experiments in *Paramecium* that inject sRNA duplexes with a 2 nt 3' overhang can target IESs for deletion, whereas the *Oxytricha* experiments shown here demonstrate that single-stranded sRNA can target IES regions for retention, suggesting an opposite mechanism. Secondly, we found that *Oxytricha*

piRNAs lack 3' modification, whereas *Tetrahymena* scnRNAs contain 3' end methylation (Kurth and Mochizuki, 2009). This difference may be part of an evolutionary mechanism leading to two orthogonally different classes of piRNAs in ciliates.

References

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res *25*, 3389-3402.

Ameres, S.L., Horwich, M.D., Hung, J.H., Xu, J., Ghildiyal, M., Weng, Z., and Zamore, P.D. (2010). Target RNA-directed trimming and tailing of small silencing RNAs. Science *328*, 1534-1539.

Aravin, A., Gaidatzis, D., Pfeffer, S., Lagos-Quintana, M., Landgraf, P., Iovino, N., Morris, P., Brownstein, M.J., Kuramochi-Miyagawa, S., Nakano, T.*, et al.* (2006). A novel class of small RNAs bind to MILI protein in mouse testes. Nature *442*, 203-207.

Ashe, A., Sapetschnig, A., Weick, E.M., Mitchell, J., Bagijn, M.P., Cording, A.C., Doebley, A.L., Goldstein, L.D., Lehrbach, N.J., Le Pen, J.*, et al.* (2012). piRNAs Can Trigger a Multigenerational Epigenetic Memory in the Germline of C. elegans. Cell *150*, 88-99.

Bartel, D.P. (2004). MicroRNAs: genomics, biogenesis, mechanism, and function. Cell *116*, 281-297.

Bouhouche, K., Gout, J.F., Kapusta, A., Betermier, M., and Meyer, E. (2011). Functional specialization of Piwi proteins in Paramecium tetraurelia from post-transcriptional gene silencing to genome remodelling. Nucleic Acids Res *39*, 4249-4264.

Brennecke, J., Aravin, A.A., Stark, A., Dus, M., Kellis, M., Sachidanandam, R., and Hannon, G.J. (2007). Discrete small RNA-generating loci as master regulators of transposon activity in Drosophila. Cell *128*, 1089-1103.

Brennecke, J., Malone, C.D., Aravin, A.A., Sachidanandam, R., Stark, A., and Hannon, G.J. (2008). An epigenetic role for maternally inherited piRNAs in transposon silencing. Science *322*, 1387-1392.

Carmell, M.A., Girard, A., van de Kant, H.J., Bourc'his, D., Bestor, T.H., de Rooij, D.G., and Hannon, G.J. (2007). MIWI2 is essential for spermatogenesis and repression of transposons in the mouse male germline. Dev Cell *12*, 503-514.

Chalker, D.L., and Yao, M.C. (1996). Non-Mendelian, heritable blocks to DNA rearrangement are induced by loading the somatic nucleus of Tetrahymena thermophila with germ line-limited DNA. Mol Cell Biol *16*, 3658-3667.

Couvillion, M.T., Lee, S.R., Hogstad, B., Malone, C.D., Tonkin, L.A., Sachidanandam, R., Hannon, G.J., and Collins, K. (2009). Sequence, biogenesis, and function of diverse small RNA classes bound to the Piwi family proteins of Tetrahymena thermophila. Genes Dev *23*, 2016-2032.

David, M., Dzamba, M., Lister, D., Ilie, L., and Brudno, M. (2011). SHRiMP2: sensitive yet practical SHort Read Mapping. Bioinformatics *27*, 1011-1012.

Duharcourt, S., Butler, A., and Meyer, E. (1995). Epigenetic self-regulation of developmental excision of an internal eliminated sequence on Paramecium tetraurelia. Genes Dev *9*, 2065-2077.

Edgar, R., Domrachev, M., and Lash, A.E. (2002). Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. Nucleic Acids Res *30*, 207-210.

Edgar, R.C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res *32*, 1792-1797.

Gan, H., Lin, X., Zhang, Z., Zhang, W., Liao, S., Wang, L., and Han, C. (2011). piRNA profiling during specific stages of mouse spermatogenesis. RNA *17*, 1191-1203.

Girard, A., Sachidanandam, R., Hannon, G.J., and Carmell, M.A. (2006). A germline-specific class of small RNAs binds mammalian Piwi proteins. Nature *442*, 199-202.
Grimson, A., Srivastava, M., Fahey, B., Woodcroft, B.J., Chiang, H.R., King, N., Degnan, B.M., Rokhsar, D.S., and Bartel, D.P. (2008). Early origins and evolution of microRNAs and Piwi-interacting RNAs in animals. Nature *455*, 1193-1197.
Grivna, S.T., Beyret, E., Wang, Z., and Lin, H. (2006). A novel class of small RNAs in mouse spermatogenic cells. Genes Dev *20*, 1709-1714.
Gu, S.G., Pak, J., Guang, S., Maniar, J.M., Kennedy, S., and Fire, A. (2012). Amplification of siRNA in Caenorhabditis elegans generates a transgenerational sequence-targeted histone H3 lysine 9 methylation footprint. Nat Genet *44*, 157-164.
Guindon, S., Dufayard, J.F., Lefort, V., Anisimova, M., Hordijk, W., and Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst Biol *59*, 307-321.
Houwing, S., Kamminga, L.M., Berezikov, E., Cronembold, D., Girard, A., van den Elst, H., Filippov, D.V., Blaser, H., Raz, E., Moens, C.B.*, et al.* (2007). A role for Piwi and piRNAs in germ cell maintenance and transposon silencing in Zebrafish. Cell *129*, 69-82.
Kirino, Y., and Mourelatos, Z. (2007). Mouse Piwi-interacting RNAs are 2'-O-methylated at their 3' termini. Nat Struct Mol Biol *14*, 347-348.
Kurth, H.M., and Mochizuki, K. (2009). 2'-O-methylation stabilizes Piwi-associated small RNAs and ensures DNA elimination in Tetrahymena. RNA *15*, 675-685.
Lau, N.C., Seto, A.G., Kim, J., Kuramochi-Miyagawa, S., Nakano, T., Bartel, D.P., and Kingston, R.E. (2006). Characterization of the piRNA complex from rat testes. Science *313*, 363-367.
Lee, H.C., Gu, W., Shirayama, M., Youngman, E., Conte, D., Jr., and Mello, C.C. (2012). C. elegans piRNAs Mediate the Genome-wide Surveillance of Germline Transcripts. Cell *150*, 78-87.
Lee, S.R., and Collins, K. (2006). Two classes of endogenous small RNAs in Tetrahymena thermophila. Genes Dev *20*, 28-33.
Lepere, G., Betermier, M., Meyer, E., and Duharcourt, S. (2008). Maternal noncoding transcripts antagonize the targeting of DNA elimination by scanRNAs in Paramecium tetraurelia. Genes Dev *22*, 1501-1512.
Lepere, G., Nowacki, M., Serrano, V., Gout, J.F., Guglielmi, G., Duharcourt, S., and Meyer, E. (2009). Silencing-associated and meiosis-specific small RNA pathways in Paramecium tetraurelia. Nucleic Acids Res *37*, 903-915.
Li, J., Yang, Z., Yu, B., Liu, J., and Chen, X. (2005). Methylation protects miRNAs and siRNAs from a 3'-end uridylation activity in Arabidopsis. Curr Biol *15*, 1501-1507.
Liu, Y., Mochizuki, K., and Gorovsky, M.A. (2004). Histone H3 lysine 9 methylation is required for DNA elimination in developing macronuclei in Tetrahymena. Proc Natl Acad Sci U S A *101*, 1679-1684.
Liu, Y., Taverna, S.D., Muratore, T.L., Shabanowitz, J., Hunt, D.F., and Allis, C.D. (2007). RNAi-dependent H3K27 methylation is required for heterochromatin formation and DNA elimination in Tetrahymena. Genes Dev *21*, 1530-1545.
Malone, C.D., and Hannon, G.J. (2009). Small RNAs as guardians of the genome. Cell *136*, 656-668.
Matzke, M., Kanno, T., Daxinger, L., Huettel, B., and Matzke, A.J. (2009). RNA-mediated chromatin-based silencing in plants. Curr Opin Cell Biol *21*, 367-376.

Mochizuki, K., Fine, N.A., Fujisawa, T., and Gorovsky, M.A. (2002). Analysis of a piwi-related gene implicates small RNAs in genome rearrangement in tetrahymena. Cell *110*, 689-699.
Mochizuki, K., and Gorovsky, M.A. (2004). Conjugation-specific small RNAs in Tetrahymena have predicted properties of scan (scn) RNAs involved in genome rearrangement. Genes Dev *18*, 2068-2073.
Mollenbeck, M., Zhou, Y., Cavalcanti, A.R., Jonsson, F., Higgins, B.P., Chang, W.J., Juranek, S., Doak, T.G., Rozenberg, G., Lipps, H.J.*, et al.* (2008). The pathway to detangle a scrambled gene. PLoS One *3*, e2330.
Nowacki, M., Higgins, B.P., Maquilan, G.M., Swart, E.C., Doak, T.G., and Landweber, L.F. (2009). A functional role for transposases in a large eukaryotic genome. Science *324*, 935-938.
Nowacki, M., Vijayan, V., Zhou, Y., Schotanus, K., Doak, T.G., and Landweber, L.F. (2008). RNA-mediated epigenetic programming of a genome-rearrangement pathway. Nature *451*, 153-158.
Ohara, T., Sakaguchi, Y., Suzuki, T., Ueda, H., and Miyauchi, K. (2007). The 3' termini of mouse Piwi-interacting RNAs are 2'-O-methylated. Nat Struct Mol Biol *14*, 349-350.
Pak, J., and Fire, A. (2007). Distinct populations of primary and secondary effectors during RNAi in C. elegans. Science *315*, 241-244.
Parfrey, L.W., Lahr, D.J., Knoll, A.H., and Katz, L.A. (2011). Estimating the timing of early eukaryotic diversification with multigene molecular clocks. Proc Natl Acad Sci U S A *108*, 13624-13629.
Patzel, V., Steidl, U., Kronenwett, R., Haas, R., and Sczakiel, G. (1999). A theoretical approach to select effective antisense oligodeoxyribonucleotides at high statistical probability. Nucleic Acids Res *27*, 4328-4334.
Prescott, D.M. (2000). Genome gymnastics: unique modes of DNA evolution and processing in ciliates. Nat Rev Genet *1*, 191-198.
Rajasethupathy, P., Antonov, I., Sheridan, R., Frey, S., Sander, C., Tuschl, T., and Kandel, E.R. (2012). A Role for Neuronal piRNAs in the Epigenetic Control of Memory-Related Synaptic Plasticity. Cell *149*, 693-707.
Robine, N., Lau, N.C., Balla, S., Jin, Z., Okamura, K., Kuramochi-Miyagawa, S., Blower, M.D., and Lai, E.C. (2009). A broadly conserved pathway generates 3'UTR-directed primary piRNAs. Curr Biol *19*, 2066-2076.
Rouget, C., Papin, C., Boureux, A., Meunier, A.C., Franco, B., Robine, N., Lai, E.C., Pelisson, A., and Simonelig, M. (2010). Maternal mRNA deadenylation and decay by the piRNA pathway in the early Drosophila embryo. Nature *467*, 1128-1132.
Saito, K., Inagaki, S., Mituyama, T., Kawamura, Y., Ono, Y., Sakota, E., Kotani, H., Asai, K., Siomi, H., and Siomi, M.C. (2009). A regulatory circuit for piwi by the large Maf gene traffic jam in Drosophila. Nature *461*, 1296-1299.
Saito, K., Nishida, K.M., Mori, T., Kawamura, Y., Miyoshi, K., Nagami, T., Siomi, H., and Siomi, M.C. (2006). Specific association of Piwi with rasiRNAs derived from retrotransposon and heterochromatic regions in the Drosophila genome. Genes Dev *20*, 2214-2222.
Saito, K., Sakaguchi, Y., Suzuki, T., Suzuki, T., Siomi, H., and Siomi, M.C. (2007). Pimet, the Drosophila homolog of HEN1, mediates 2'-O-methylation of Piwi- interacting RNAs at their 3' ends. Genes Dev *21*, 1603-1608.

Schoeberl, U.E., Kurth, H.M., Noto, T., and Mochizuki, K. (2012). Biased transcription and selective degradation of small RNAs shape the pattern of DNA elimination in Tetrahymena. Genes Dev *26*, 1729-1742.
Seto, A.G., Kingston, R.E., and Lau, N.C. (2007). The coming of age for Piwi proteins. Mol Cell *26*, 603-609.
Shirayama, M., Seth, M., Lee, H.C., Gu, W., Ishidate, T., Conte, D., Jr., and Mello, C.C. (2012). piRNAs Initiate an Epigenetic Memory of Nonself RNA in the C. elegans Germline. Cell *150*, 65-77.
Sijen, T., Steiner, F.A., Thijssen, K.L., and Plasterk, R.H. (2007). Secondary siRNAs result from unprimed RNA synthesis and form a distinct class. Science *315*, 244-247.
Storici, F. (2008). RNA-mediated DNA modifications and RNA-templated DNA repair. Curr Opin Mol Ther *10*, 224-230.
Tian, Y., Simanshu, D.K., Ma, J.B., and Patel, D.J. (2011). Structural basis for piRNA 2'-O-methylated 3'-end recognition by Piwi PAZ (Piwi/Argonaute/Zwille) domains. Proc Natl Acad Sci U S A *108*, 903-910.
Vagin, V.V., Sigova, A., Li, C., Seitz, H., Gvozdev, V., and Zamore, P.D. (2006). A distinct small RNA pathway silences selfish genetic elements in the germline. Science *313*, 320-324.
Verdel, A., Jia, S., Gerber, S., Sugiyama, T., Gygi, S., Grewal, S.I., and Moazed, D. (2004). RNAi-mediated targeting of heterochromatin by the RITS complex. Science *303*, 672-676.
Volpe, T.A., Kidner, C., Hall, I.M., Teng, G., Grewal, S.I., and Martienssen, R.A. (2002). Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi. Science *297*, 1833-1837.
Wang, J., Czech, B., Crunk, A., Wallace, A., Mitreva, M., Hannon, G.J., and Davis, R.E. (2011). Deep small RNA sequencing from the nematode Ascaris reveals conservation, functional diversification, and novel developmental profiles. Genome Res *21*, 1462-1477.
Wassenegger, M., Heimes, S., Riedel, L., and Sanger, H.L. (1994). RNA-directed de novo methylation of genomic sequences in plants. Cell *76*, 567-576.
Watanabe, T., Takeda, A., Tsukiyama, T., Mise, K., Okuno, T., Sasaki, H., Minami, N., and Imai, H. (2006). Identification and characterization of two novel classes of small RNAs in the mouse germline: retrotransposon-derived siRNAs in oocytes and germline small RNAs in testes. Genes Dev *20*, 1732-1743.
Watanabe, T., Tomizawa, S., Mitsuya, K., Totoki, Y., Yamamoto, Y., Kuramochi-Miyagawa, S., Iida, N., Hoki, Y., Murphy, P.J., Toyoda, A.*, et al.* (2011). Role for piRNAs and noncoding RNA in de novo DNA methylation of the imprinted mouse Rasgrf1 locus. Science *332*, 848-852.
Wei, W., Ba, Z., Gao, M., Wu, Y., Ma, Y., Amiard, S., White, C.I., Danielsen, J.M., Yang, Y.G., and Qi, Y. (2012). A Role for Small RNAs in DNA Double-Strand Break Repair. Cell *149*, 101-112.
Witherspoon, D.J., Doak, T.G., Williams, K.R., Seegmiller, A., Seger, J., and Herrick, G. (1997). Selection on the protein-coding genes of the TBE1 family of transposable elements in the ciliates Oxytricha fallax and O. trifallax. Mol Biol Evol *14*, 696-706.
Yao, M.C., Fuller, P., and Xi, X. (2003). Programmed DNA deletion as an RNA-guided system of genome defense. Science *300*, 1581-1584.
Zahler, A.M., Neeb, Z.T., Lin, A., and Katzman, S. (2012). Mating of the Stichotrichous Ciliate Oxytricha trifallax Induces Production of a Class of 27 nt Small RNAs Derived from the Parental Macronucleus. PLoS One *7*, e42371.

Zamore, P.D., and Haley, B. (2005). Ribo-gnome: the big world of small RNAs. Science *309*, 1519-1524.
Zhang, H., Ehrenkaufer, G.M., Pompey, J.M., Hackney, J.A., and Singh, U. (2008). Small RNAs with 5'-polyphosphate termini associate with a Piwi-related protein and regulate gene expression in the single-celled eukaryote Entamoeba histolytica. PLoS Pathog *4*, e1000219.

Experimental Procedure

Cell culture and mating

Oxytricha trifallax strains JRB310 and JRB510 were cultured in Pringsheim medium (0.11 mM Na_2HPO_4, 0.08 mM $MgSO_4$, 0.85 mM $Ca(NO_3)_2$, 0.35 mM KCl, pH 7.0) using algae *Chlamydomonas reinhardtii* as a food source. To induce mating, JRB310 and LRB510 cells were mixed right after food depletion. Cells started to pair between 2-3 hr post mixing, and conjugation efficiency was between 60% and 95%. Post-injection cells were raised in Volvic water.

RNA extraction and small RNA detection

Total RNA was extracted using the mirVana kit (Ambion) following manufacturer's instructions. To detect small RNA, total RNA was separated on a 15% denaturing polyacrylamide gel in 1× TBE buffer and stained with SYBR Gold (Invitrogen). For radioactive labeling, RNA was treated with calf intestinal alkaline phosphatase (NEB), phenol-chloroform extracted, ethanol precipitated, and labeled with γ-^{32}P-ATP using T4 polynucleotide kinase (NEB). Labeled RNA was separated on a 15% denaturing polyacrylamide gel, which was directly exposed to a storage phosphor screen (GE healthcare) and scanned by Typhoon 9410 (GE healthcare).

Phylogenetic analysis of *Oxytricha* Argonaute proteins

Oxytricha Argonaute genes (*Otiwi1-13*) were retrieved by using tBLASTn (Altschul et al., 1997) against the current macronuclear genome assembly (Swart et al., in revision) and *Tetrahymena* TWI1p (GenBank: BAC02573.1) as the query sequence. RT-PCR followed by clone sequencing was used to generate cDNA and protein sequences. All thirteen Otiwi proteins, selected eukaryotic (accession numbers from (Seto et al., 2007)) and archaeal (accession numbers NP_248321 and NP_578266) Argonaute proteins were aligned by MUSCLE (Edgar, 2004) followed by manual adjustment, which included deletion of multiple gap positions (see the sequence alignment in a separate supplemental file labeled “Argonaute_alignment”). Phylogenetic analysis was performed using maximum likelihood through the PhyML webserver (Guindon et al., 2010) with default settings and the LG substitution matrix. Bootstrap values were produced by 500 re-sampled datasets.

Antibody

The antibody used to detect Otiwi1 in western analysis, immunostaining, and immunoprecipitation is anti-PIWIL1 (Abcam, ab12337), originally raised against the human PIWIL1 protein C-terminal peptide (15aa). We performed knockdown followed by western analysis, peptide competition in immunostaining, and mass spectrometry after immunoprecipitation to verify that the antibody is specific to Otiwi1. Detailed protocols for immunostaining, peptide competition assay, immunoprecipitation, and mass spectrometry analysis are provided in Extended Experimental Procedures. Anti-tubulin antibody is DM1A (Abcam, ab7291).

Biochemical analysis of piRNAs
For detecting a 5' monophosphate, ~ 40ng Otiwi1 IP-enriched RNA (23 h) or control RNA oligonucleotide was incubated with 0.1U Terminator exonuclease (Epicentre) at 30°C for 1h, and the reaction was quenched by adding EDTA (pH 8.0) to a final concentration of 5 mM. The quenched reaction was directly loaded and separated on a 15% denaturing polyacrylamide gel, and stained with SYBR gold (Invitrogen). To analyze 3' end modifications, *Oxytricha* piRNAs were treated with $NaIO_4$, followed by β-elimination as described previously (Vagin et al., 2006).

Antisense knockdown
To increase the efficiency of targeting, mRNA secondary structures were used to guide antisense oligonucleotide design (Patzel et al., 1999). The oligonucleotides, at a final concentration of 30 µg/µl, were injected into cytoplasm right after pair formation (~2-4 hr post mixing). To verify the knockdown effect, total RNA was extracted at 10-11 hr post-mixing for RT-qPCR analysis (Figure S3A). We also collected cells at 10 hr and 18 hr post-mixing for western analysis after Otiwi1 knockdown (Figure S3B). See Extended Experimental Procedures for verification of knockdown by RT-PCR and western analysis.

Survival and developmental characterization of *Otiwi1* knockdown
After antisense injection, 20-28 pairs were collected into one droplet (0.2% BSA in Volvic water) on a cover slip, covered with oil. We then observe and counted cells using light microscopy at 12, 24, 28, and 48 hr post mixing. Student's t-test was used to compare the number of rounded cells at 24 hr based on four independent knockdown injections. In Figures 3C-J, single cells at 24-28 hr post mixing were isolated into droplets, fixed with Methanol:Acetic acid (3:1), stained with DAPI, and imaged with laser scanning confocal microscopy at the Princeton University Microscopy Facility.

sRNA Illumina library construction, high throughput sequencing and analysis
Otiwi1 antibody immunoprecipitated RNA from 12, 19, 23, and 30 hr post mixing, as well as total sRNA isolated at 19 hr after backcrossing the Contig22226.0 $IES1^+$ strain (a clonal line derived from an F1 cell from the population analyzed in Figure 5A lane 1) to wild-type JRB510, were used to construct sRNA libraries, following the Illumina® TruSeq™ Small RNA Sample Preparation protocol, with different barcodes for each sample. The libraries were pooled and sequenced on an Illumina HiSeq 2000 (101 cycles, single-end reads, multiplexed) at the Sequencing Core Facility at Princeton University. The raw reads were barcode-split, adapter trimmed, and collapsed to non-redundant reads before mapping by SHRiMP (David et al., 2011). The total sRNA library at 20 hr post-mixing was prepared following the Illumina® Small RNA v1.5 Sample Preparation guide, and sequenced on an Illumina Genome Analyzer IIX (54 cycles, single-end reads). See Extended Experimental Procedures for details of sequencing data processing and analysis.

Synthetic RNA and DNA oligonucleotide injections for IES retention experiments
RNA and DNA oligonucleotides were synthesized by IDT with standard desalting. Oligonucleotides were dissolved in nuclease-free water (Ambion) to a final concentration of 20 µg/µl, heated to 65 °C for 1 min, and chilled on ice before using. Paired cells between 12 to 18 hr post mixing were isolated in single droplets on cover slips and covered with mineral oil during injection. For each oligonucleotide, the cytoplasm of both cells from 20-40 pairs were injected and all pairs pooled together at the end of injection. Cells were examined for donut shape

between 36 to 48 hr post mixing and morphologically vegetative cells were removed. The pooled injected cells gave rise to a starting population of 5-20 cells 3-4 days post-mixing when they started vegetative division. At 7-8 days post-mixing when genomic DNA was extracted, the final population varied between 50 to 300 cells depending on the cell conditions. DNA equivalent of that from 0.5-3 cells was used in each PCR reaction. For single-cell analysis, single cells were hand-isolated in 1 µl medium from the injected population on Day 7, and directly used in PCR.

For F2 analysis, F1 cells were cultured vegetatively for 10 ~ 20 days (or 10 ~ 40 generations) in pools that we described above. We then starved them to induce conjugation. Conjugating pairs were hand-isolated and pooled in a separate well. We allowed the conjugation to proceed and vegetative growth for at least 5 days (5~10 generations) before genomic DNA extraction. When we could not induce F1 cells to mate with each other, we then tried to induce conjugation between F1 cells and the WT JRB310 or JRB510 parental strain, which usually gave higher mating efficiency. Contig 22226.0 IES1$^+$ F2 and F3 derive from a mating between individual F1 and F2 cell lines, respectively.

Accession numbers

Otiwi protein sequences are available with GenBank accession numbers JN604928-JN604940. The raw and processed sRNA sequencing data have been deposited in NCBI's Gene Expression Omnibus (Edgar et al., 2002) with the GEO accession numbers GSE35018 and GSE40081.

Acknowledgements

We thank D. Perlman, B. Zee, and B. Garcia for mass spectrometry validation of the Otiwi1 antibody; J. Buckles, W. Wang, D. Storton, and L. Parsons for Illumina sequencing; L. Li and A. D. Goldman for assistance with phylogenetic analysis; X. Chen for searching the micronuclear genome assembly for junction-spanning piRNAs; A. Chen, K. Mochizuki and M. Couvillion for sharing RNA immunoprecipitation protocols; E. Swart, J. Bloom and P. Jiang for advice on computational analysis of deep sequencing results; J. Postberg and J. Goodhouse for immunofluorescence advice; P. Andolfatto, J. Swan, J. Khurana, P. Schedl and L. Beh for discussion or comments on the manuscript; J. Wang for help with *Oxytricha* cell culture. We also thank three anonymous referees for valuable suggestions. This study was supported by NIH grant GM59708 and NSF grants 0923810 and 0900544 (to L.F.L.) and a DOD pre-doctoral fellowship W81XWH-10-1-0122 (to W.F.).

Figure Legends

Figure 1. 27 nt sRNAs and Otiwi1 co-express and interact during *Oxytricha* conjugation. (A) 27 nt sRNAs (filled arrowhead) are abundant in early conjugation. Total RNA extracted at different time points during conjugation was separated on a denaturing polyacrylamide gel and stained with SYBR gold. *Chlamydomonas* (food algae) RNA is also loaded as a control. (B) A class of ~21 nt sRNAs (open arrowhead) is present in weaker abundance in both vegetative and conjugating *Oxytricha* cells. (C) Phylogenetic analysis of Piwi proteins suggests that Otiwi proteins form two major clades. Archaeal Argonaute proteins were used as outgroups for eukaryotic proteins. Otiwi clade I (Otiwi1-4 and 11) members are labeled orange, and Piwi subfamily members from animals and *Dictyostelium* are labeled red. Otiwi1, which associates with 27 nt sRNAs in *Oxytricha,* is indicated by a star. Green: Otiwi clade II (Otiwi5-10, 12, and 13), whose precise phylogeny is not resolved with high confidence (low bootstrap value on relevant nodes); blue: Ago subfamily of Argonaute proteins from plants, fungi and animals.

Bootstrap values greater than 60% are shown. Abbreviations: At: *Arabidopsis thaliana*; Dd: *Dictyostelium discodium*; Dm: *Drosophila melanogaster*; Dr: *Danio rerio*; Mj: *Methanocaldococcus jannaschii*; Mm: *Mus musculus*; Nc: *Neurospora crassa*; Pf: *Pyrococcus furiosus*; Sp: *Schizosaccaromyces pombe*. Twi and Ptiwi are Argonaute proteins from *Tetrahymena thermophila* and *Paramecium tetraurelia*, respectively. The scale bar represents 0.5 substitution per site (inferred by LG model). (D) *Otiwi* genes differ in expression profile across *Oxytricha*'s life cycle. The orange labels indicate *Otiwi* clade I that clusters with the Piwi subfamily in (C), and these *Otiwi* genes are strongly induced during conjugation (post mixing). The green labels indicate *Otiwi* clade II, whose mRNA expression is more uniform. *Actin I* RT-PCR is used as a loading control. V, vegetative cells; s, starved cells; RT, reverse transcription. (E) Otiwi1 immunoprecipitation enriches for 27 nt sRNAs from total RNA at 12, 19, 23, and 30 hr post mixing of two mating types. (F) Northern analysis of *Otiwi1* expression from the conjugation time series shown in (A) and (B). (G) Western analysis of *Otiwi1* expression from the conjugation time series shown in (A) and (B). (H) Terminator treatment indicates that *Oxytricha* piRNAs (O.t. piRNAs) contain a 5' monophosphate. Control oligonucleotides are 27 nt synthetic RNAs with either a 5' monophosphate or a 5'-OH. (I) Periodate oxidation followed by β-elimination suggests that *Oxytricha* piRNAs (O.t. piRNAs) lack 3' end modification. The control oligonucleotide is a 27 nt synthetic RNA with 2'-O methylation at the 3' end. (J) Otiwi1 lacks a Piwi-specific insertion (red box) predicted to help accommodate the 2'-O-methylated 3' ends of mammalian piRNAs (Tian et al., 2011). See also Figure S1 and Table S1.

Figure 2. Otiwi1 localization shifts from old macronucleus to the developing macronucleus. Otiwi1 is absent at the initial pairing stage when the micronucleus (MIC) begins meiosis (A), but accumulates in the old macronucleus (MAC) during mid-to-late pair formation (B). Otiwi1 is transiently present in both the cytoplasm and the developing macronucleus (called anlagen, AN) right after its formation (C), and localizes only to the anlagen in early exconjugants (D). (E) Otiwi1 signal decreases dramatically in late anlagen development, and the parental MAC is degrading. Arrowhead in (A) indicates a MIC undergoing meiosis and stained with α-tubulin. Green, Otiwi1; red, 4', 6-diamidino-2-phenylindole (DAPI); white, α-tubulin. See also Figure S2.

Figure 3. *Otiwi1* knockdown (KD) leads to a depletion of 27 nt piRNAs and stalled nuclear development. (A) 27 nt piRNAs are lost upon *Otiwi1* but not *Otiwi4* KD. *Otiwi1* or *Otiwi4* antisense, or control oligonucleotides were injected at 2-4 hr post-mixing, and total RNA was extracted 18 hr post-mixing for radioactive labeling and detection. (B) Cells were arrested at pair stage or shortly after pair separation upon *Otiwi1* KD, and died between 24 to 48 hr post-mixing. The quantification of survived cells was based on four independent experiments injecting 20-28 pairs each, for both KD and control. We calculate the percentage of cells based on the starting number of cells, not the population at indicated time points. Error bars indicate standard error of the mean. (C-J) Representative microscopy images of control and KD cells at 24-28 hr post-mixing. Arrested pairs show either nuclear morphology resembling wild-type cells at 12h (C), or degenerated nuclei (D). (E, F) Most control cells show normal donut-shaped morphology, where the round, light-colored "hole" with faint DAPI staining is the developing nucleus, or anlagen (AN). (G-J) Significantly more *Otiwi1* KD cells show abnormal spherical structures in which any one or all of the macronucleus (MAC), micronucleus (MIC) and developing macronulceus (AN) are missing. Scale bars are 14 μm in (C) and (E), and 7μm in (D, F, G-J). See also Figure S3.

Figure 4. Most *Oxytricha* piRNAs map to sequences that are retained during genome rearrangement. (A) The length distribution of *Oxytricha* piRNAs suggests that they are predominantly 27 nt. Otiwi1 immunoprecipitated sRNA at 19 hr and total sRNA at 20 hr are shown. 19 and 20 hr conjugation populations do not differ significantly regarding staging. (B) Sequence profile of *Oxytricha* 27 nt sRNA reads from either Otiwi1 immunoprecipitation at 19 hr or total sRNA at 20 hr. (C) Mapping statistics of 27 nt, 5'-U Otiwi1-associated piRNAs at 12, 19, 23, and 30 hr post-mixing. Reads mapping to the *Oxytricha* mitochondria or *Chlamydomonas* (food algae) genomes serve as negative controls. (D) Mapping of 27 nt, 5'-U piRNAs to macronuclear and micronuclear loci for *TEBPβ, actin I* (both 19 hr) and *Otiwi3* (12 hr). (E) Metagene analysis (N = 10,497) suggests that piRNAs distribute across the entire macronuclear chromosomes. For each mapping, the 13 nt position of a piRNA read is normalized against the length of the corresponding single-gene macronuclear chromosome, with red indicating the mRNA sense strand for protein-coding genes and blue indicating antisense strand. (F) Genome-wide nucleotide representation by piRNA reads suggests that the whole macronuclear genome is converted into piRNAs. For each macronuclear chromosome surveyed, the nucleotide coverage is calculated as piRNA-covered sequence length divided by the contig length, and the distribution of nucleotide representation is plotted for all 10,497 contigs examined. (G) Genome-wide nucleotide representation by piRNA reads mapping to either sense (red) or antisense (blue) strand of protein-coding genes, as calculated in (F). Only 27 nt 5'-U reads from 19 hr IP were shown for (E), (F) and (G), but the pattern holds for all four time-points. See also Table S2, Figure S4 and Figure S5.

Figure 5. Small RNA injection leads to heritable retention of normally-deleted genomic regions. PCR amplification of total DNA from exconjugant cells after corresponding synthetic sRNA injection reveals programmed IES retention (labeled IES^+ MAC). (A) PCR assay for the presence of a larger IES^+ MAC PCR product if the targeted conventional IES between MDS1-2 is retained in MAC Contig22226.0. Bands corresponding to wild-type (WT) MAC PCR products are also indicated. (B) PCR assays for a smaller *actin I* MAC PCR product if the scrambled IES between MDS5-7 is retained instead of being replaced by the longer, scrambled MDS6. (C) Selective PCR assay for the presence of a smaller *actin I* MAC PCR product if the targeted IES between MDS7-9 is retained; larger bands correspond to the scrambled precursor MIC DNA. (D) PCR assay for the presence of a smaller *actin I* MAC PCR product if the scrambled IES between MDS10-12 is retained. MIC contig structure, synthetic oligonucleotides injected (purple), and PCR products are shown schematically on the right. MDSs (white boxes with numbers) and IESs (grey and red boxes) are not drawn to scale. Red boxes represent IESs that are targeted by injected RNAs, and black boxes denote short telomeres on nanochromosomes. Arrows schematically indicate primer positions. (E) Exconjugant cells from a backcross between the Contig22226.0 $IES1^+$ strain and a parental wild-type strain ($IES^+ \times$ WT) produce new IES-containing piRNAs, but wild-type mating cells (WT × WT) do not. 25-28 nt 5'-U reads are plotted. Forward and reverse mappings are displayed above and below the relevant portion of the full-length annotated contig, respectively. Stars indicate two new IES-containing piRNAs. Gaps indicate WT reads mapping across the MDS6-7 junction. See also Figure S6 and sequence alignments of all PCR products in a supplemental file.

Figure 6. Model: piRNA protects DNA against loss during *Oxytricha* genome rearrangement. During conjugation, the parental macronuclear genome produces long non-coding RNAs which are subsequently processed into 27 nt piRNAs. Otiwi1 loads 27 nt piRNAs

and transport them into the developing macronucleus. The Otiwi1-piRNA complex recognizes and helps mark MDS regions for retention. TBE transposase proteins may introduce DNA breaks in IES regions that are then eliminated, and MDS segments fuse to form new macronuclear chromosomes.

Figure 1

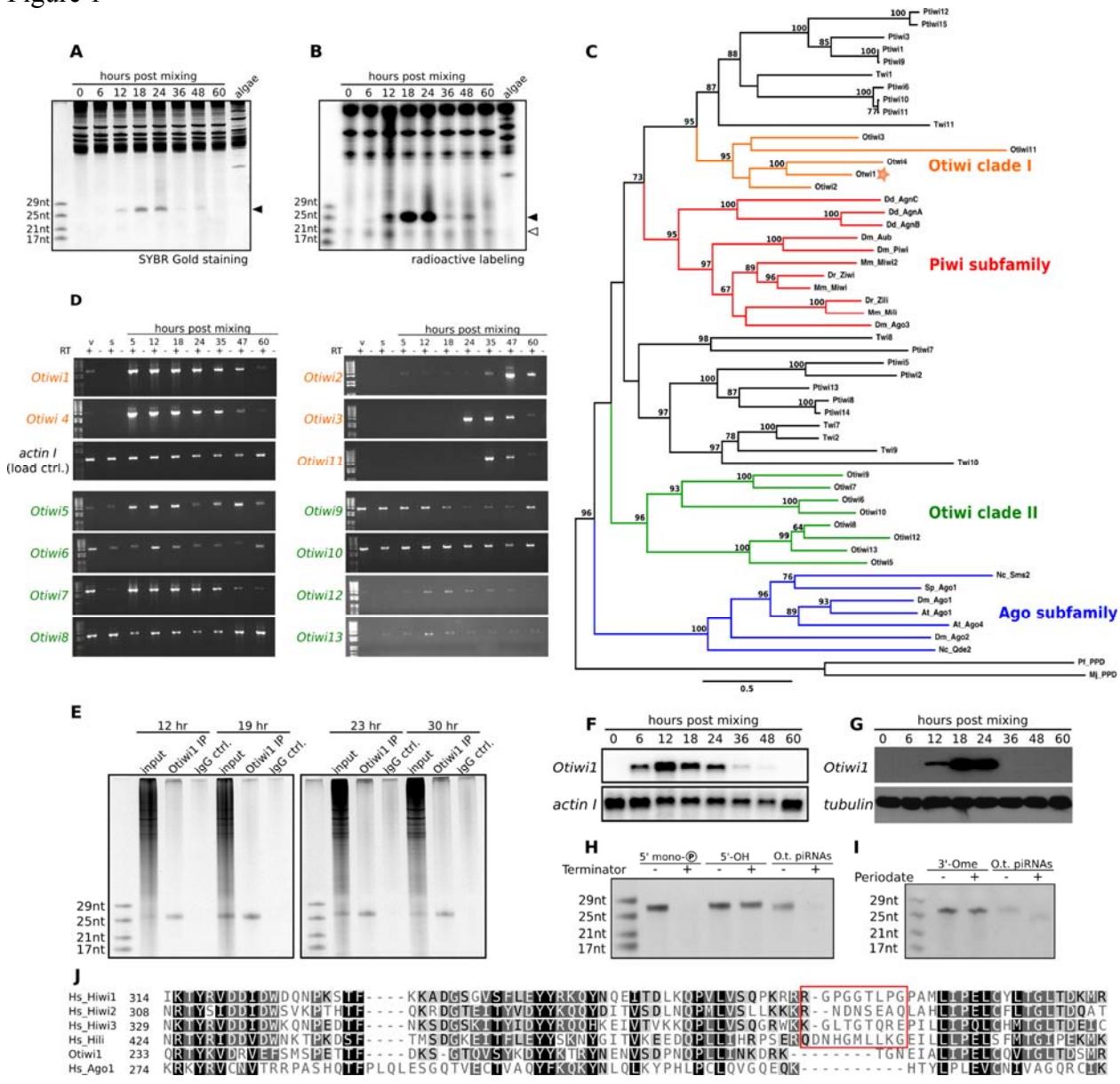

A
hours post mixing
0 6 12 18 24 36 48 60 algae
29nt
25nt
21nt
17nt
SYBR Gold staining
B
hours post mixing
0 6 12 18 24 36 48 60 algae
radioactive labeling
C
Otiwi clade I
Piwi subfamily
Otiwi clade II
Ago subfamily
0.5
D
hours post mixing
RT
Otiwi1
Otiwi 4
actin I
(load ctrl.)
Otiwi5
Otiwi6
Otiwi7
Otiwi8
Otiwi2
Otiwi3
Otiwi11
Otiwi9
Otiwi10
Otiwi12
Otiwi13
E
12 hr
19 hr
23 hr
30 hr
input
Otiwi1 IP
IgG ctrl.
29nt
25nt
21nt
17nt
F
hours post mixing
0 6 12 18 24 36 48 60
Otiwi1
actin I
G
hours post mixing
0 6 12 18 24 36 48 60
Otiwi1
tubulin
H
Terminator
5' mono-P
5'-OH
O.t. piRNAs
29nt
25nt
21nt
17nt
I
Periodate
3'-Ome
O.t. piRNAs
29nt
25nt
21nt
17nt
J
Hs_Hiwi1 314
Hs_Hiwi2 308
Hs_Hiwi3 329
Hs_Hili 424
Otiwi1 233
Hs_Ago1 274

Figure 2

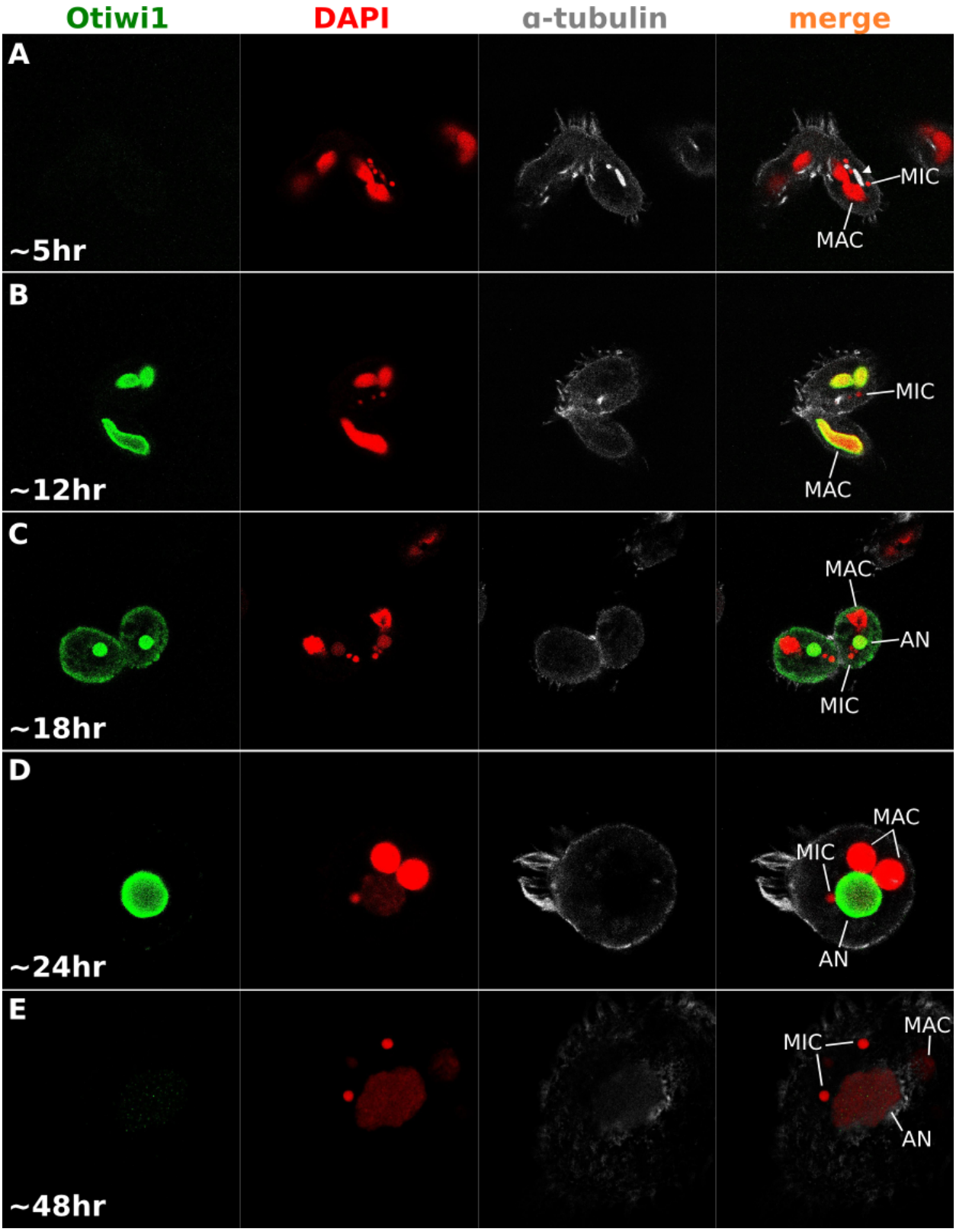

Otiwi1
DAPI
α-tubulin
merge
A
~5hr
MIC
MAC
B
~12hr
MIC
MAC
C
~18hr
MAC
AN
MIC
D
~24hr
MAC
MIC
AN
E
~48hr
MIC
MAC
AN

Figure 3

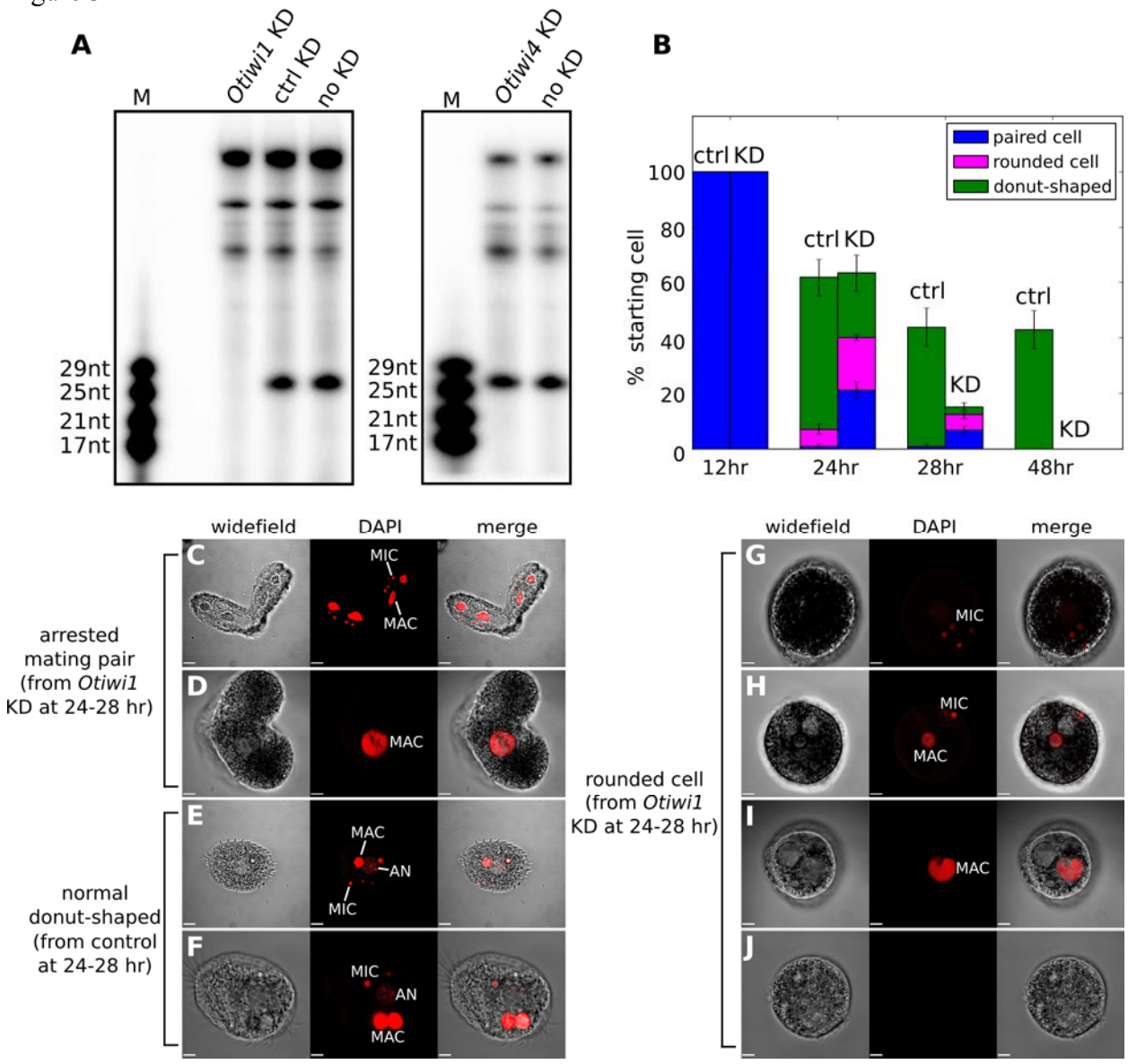

A
M
Otiwi1 KD
ctrl KD
no KD
29nt
25nt
21nt
17nt
M
Otiwi4 KD
no KD
29nt
25nt
21nt
17nt
B
paired cell
rounded cell
donut-shaped
% starting cell
ctrl KD
ctrl KD
ctrl
KD
ctrl
KD
12hr
24hr
28hr
48hr
widefield
DAPI
merge
arrested mating pair (from Otiwi1 KD at 24-28 hr)
normal donut-shaped (from control at 24-28 hr)
rounded cell (from Otiwi1 KD at 24-28 hr)
MIC
MAC
AN

Figure 4

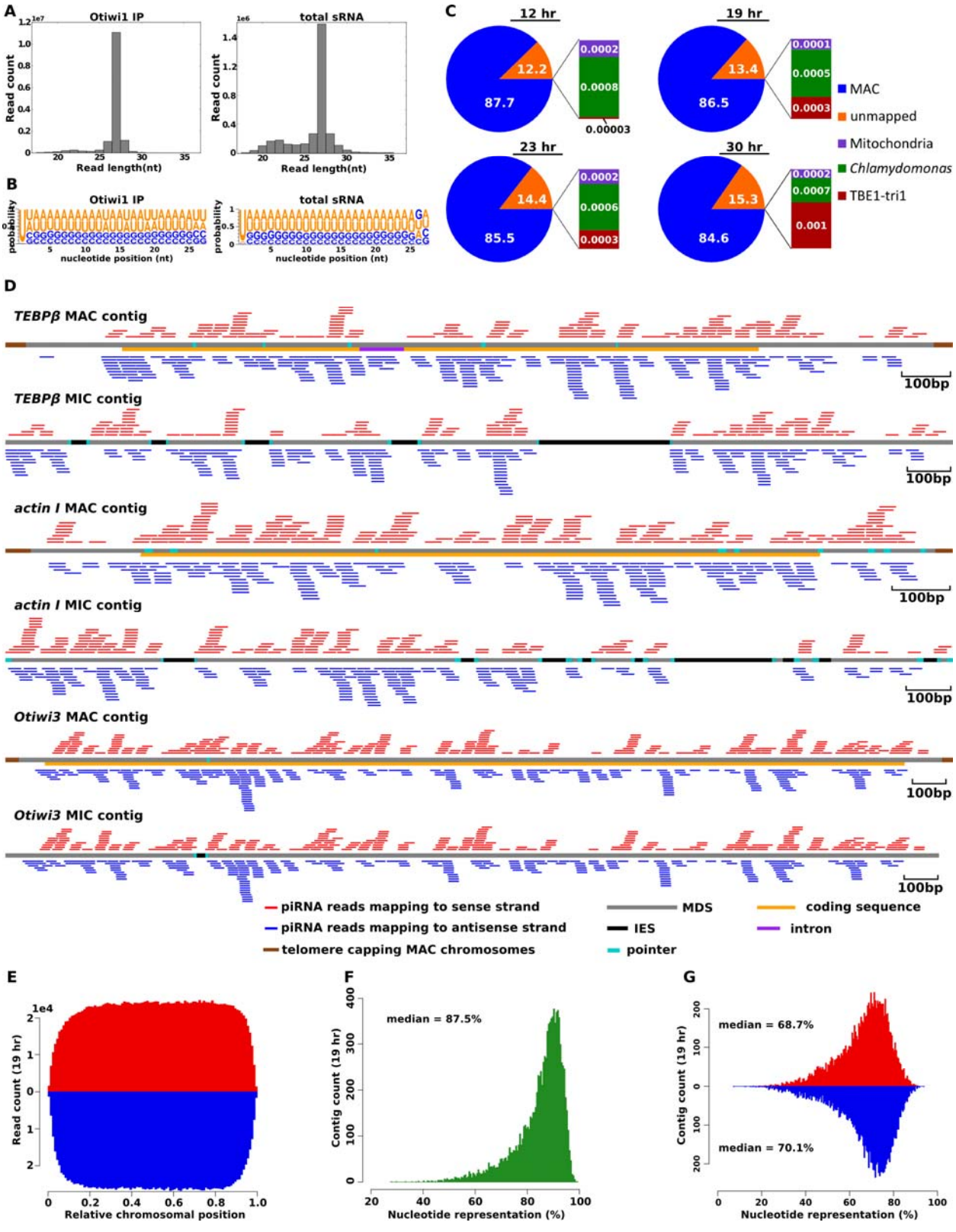

A
Otiwi1 IP
total sRNA
Read count
Read length(nt)
B
Otiwi1 IP
total sRNA
probability
nucleotide position (nt)
C
12 hr
19 hr
23 hr
30 hr
12.2
87.7
13.4
86.5
14.4
85.5
15.3
84.6
0.0002
0.0008
0.00003
0.0001
0.0005
0.0003
0.0006
0.0007
0.001
MAC
unmapped
Mitochondria
Chlamydomonas
TBE1-tri1
D
TEBPβ MAC contig
TEBPβ MIC contig
actin I MAC contig
actin I MIC contig
Otiwi3 MAC contig
Otiwi3 MIC contig
100bp
piRNA reads mapping to sense strand
piRNA reads mapping to antisense strand
telomere capping MAC chromosomes
MDS
IES
pointer
coding sequence
intron
E
Read count (19 hr)
Relative chromosomal position
F
Contig count (19 hr)
median = 87.5%
Nucleotide representation (%)
G
median = 68.7%
median = 70.1%
Contig count (19 hr)
Nucleotide representation (%)

Figure 5

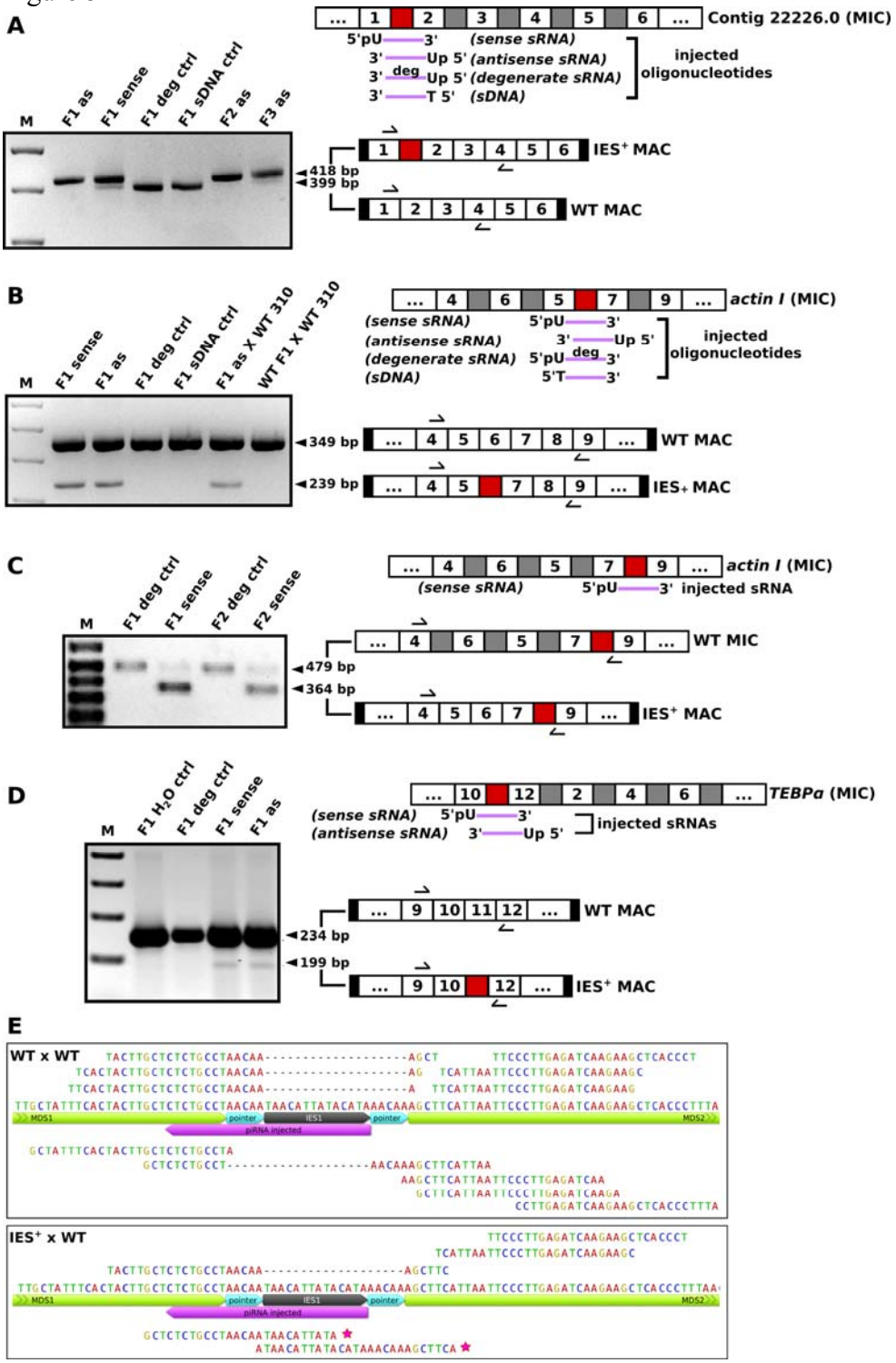
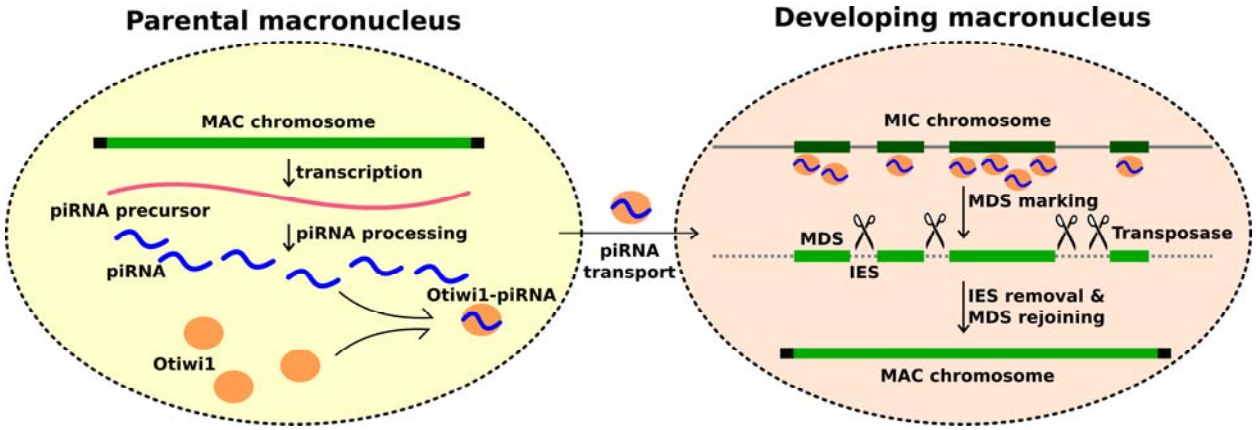

A
M
F1 as
F1 sense
F1 deg ctrl
F1 sDNA ctrl
F2 as
F3 as
418 bp
399 bp
Contig 22226.0 (MIC)
5'pU 3' (sense sRNA)
3' Up 5' (antisense sRNA)
3' deg Up 5' (degenerate sRNA)
3' T 5' (sDNA)
injected oligonucleotides
IES+ MAC
WT MAC
B
M
F1 sense
F1 as
F1 deg ctrl
F1 sDNA ctrl
F1 as X WT 310
WT F1 X WT 310
349 bp
239 bp
actin I (MIC)
(sense sRNA) 5'pU 3'
(antisense sRNA) 3' Up 5'
(degenerate sRNA) 5'pU deg 3'
(sDNA) 5'T 3'
injected oligonucleotides
WT MAC
IES+ MAC
C
M
F1 deg ctrl
F1 sense
F2 deg ctrl
F2 sense
479 bp
364 bp
actin I (MIC)
(sense sRNA) 5'pU 3' injected sRNA
WT MIC
IES+ MAC
D
M
F1 H2O ctrl
F1 deg ctrl
F1 sense
F1 as
234 bp
199 bp
TEBPα (MIC)
(sense sRNA) 5'pU 3'
(antisense sRNA) 3' Up 5'
injected sRNAs
WT MAC
IES+ MAC
E
WT x WT
MDS1
pointer
IES1
pointer
MDS2
piRNA injected
IES+ x WT
MDS1
pointer
IES1
pointer
MDS2
piRNA injected

Figure 6


Parental macronucleus
MAC chromosome
transcription
piRNA precursor
piRNA processing
piRNA
Otiwi1-piRNA
Otiwi1
piRNA
transport
Developing macronucleus
MIC chromosome
MDS marking
MDS
IES
Transposase
IES removal &
MDS rejoining
MAC chromosome

Supplemental Information for

Piwi-Interacting RNAs Protect DNA Against Loss During *Oxytricha* Genome Rearrangement

Wenwen Fang, Xing Wang, John R. Bracht, Mariusz Nowacki, Laura F. Landweber*

*To whom correspondence should be addressed.
E-mail: lfl@princeton.edu

This file includes:

Supplemental Figures

Extended Experimental Procedures

Oligonucleotide Sequences

Supplemental References

Tables S1 and S2

Supplemental Figure Legends

Figure S1. Mass spectrometry verification of Otiwi1 antibody, related to Figure 1. (A) Immunoprecipitation (IP) with anti-PIWIL1 antibody at 12, 19, 23, and 30 hr post-mixing enriched for a ~90 kD protein band which was sequenced. The ~50 kD protein is rabbit IgG heavy chain which we also sequenced by mass spectrometry (data not shown). Otiwi1 antibody IP (IP), IgG control IP (IgG ctrl.) and input samples from all four time points in the crosslinking RNA co-IP experiments were separated by SDS-PAGE and stained with GelCode Blue Stain Reagent. (B) Sequence coverage of *Otiwi1* at 23 hr post-mixing. Red letters show the resulting peptide coverage from mass spectrometry sequencing. (C) Representative MS/MS of $[M+2H]^{2+}$ ion of a Otiwi1 peptide from the 23 hr post-mixing sample.

Figure S2. *Oxytricha* sexual cycle and peptide competition assay for Otiwi1 antibody, related to Figure 2. (A) *Oxytricha* sexual cycle illustrated by immunofluorescence microscopy. Each vegetative *Oxytricha* cell contains two macronuclei (MAC) and two micronuclei (MIC). Upon conjugation, two cells of opposite mating type pair, the MAC elongates and MIC undergoes meiosis. One haploid MIC exchanges between two cells, and fuses with its partner to form a zygote. The diploid zygotic nucleus divides mitotically twice. One of the mitotic products differentiates into the new MAC, one is destroyed, and two become the new MIC. Anlagen (AN) is the nucleus where genome rearrangement occurs, events that involve chromosome polytenization, MIC-limited DNA deletion and gene unscrambling. These processes are accompanied by an increase in AN volume, and the disintegration of the old MAC. Chromosome fragmentation, telomere addition, and a second amplification produce "nanochromosomes" in the mature new MAC. Green, H3K4me3 antibody staining to indicate "active" chromatin; red, DAPI; white, α-tubulin. (B) Peptide competition assay indicates that the Otiwi1 antibody has a much higher affinity for C-terminus of Otiwi1 compared to Otiwi4. Representative images are cells immunostained with Otiwi1 antibody with indicated peptide competition. Scale bars represent 9μm and 11μm in pair stage and early donut-shaped cells, respectively. Below each representative images are zoomed-out images of immunostaining shown above. The viewed population is a mixture of vegetative cells, paired cells, and early exconjugants. Only mid-to-late pairs and early exconjugants show signal from Otiwi1 antibody, but DAPI stains all cells. Scale bar represents 260 μm. Green, Otiwi1 antibody; red, DAPI; white, α-tubulin.

Figure S3. Verification of antisense knockdown (KD), related to Figure 3. (A) RT-qPCR analysis of *Otiwi1* and *Otiwi4* KD. *Otiwi1* mRNA level is ~100-fold reduced upon *Otiwi1* KD, compared to control-injection cells, whereas *Otiwi4* mRNA level is comparable to control. *Otiwi4* mRNA level is ~6.4 fold reduced upon *Otiwi4* KD, compared to control cells, whereas *Otiwi1* mRNA level is comparable to control**.** mRNA levels of *Otiwi1* and *Otiwi4* at 10-11 hr post-mixing were assessed by qPCR, normalized against mitochondrial rRNA. Due to limited quantities of material harvested from injected cells, we could not detect reliable signals from *Otiwi3* and *Otiwi8* transcripts, but we reason that their divergence from *Otiwi1* and *Otiwi4* at the nucleotide level should preclude a cross-knockdown effect. (B) Loss of Otiwi1 antibody signal in western analysis upon *Otiwi1* antisense KD.

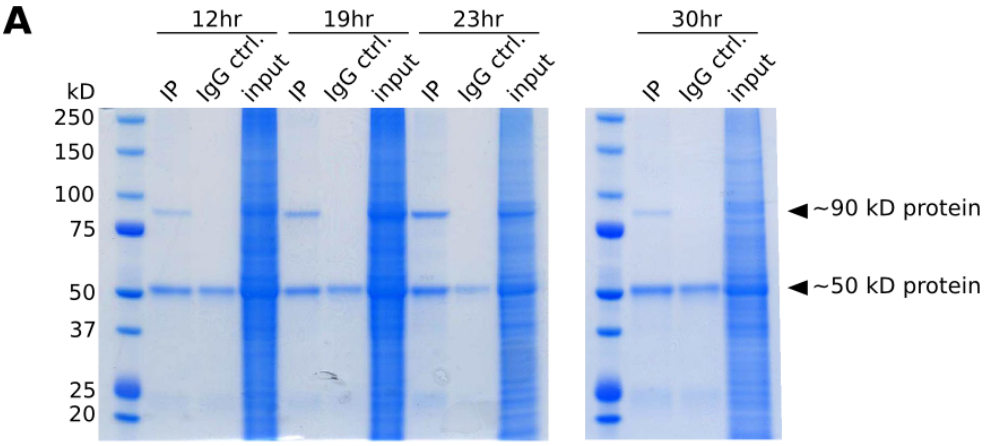
Figure S4. piRNAs mapping to the *TEBPβ* MAC and MIC contig at different time points show a similar pattern, related to Figure 4. Only 27nt 5'-U reads were plotted. The similar peak pattern in different samples suggests a bias in either biogenesis or the Illumina-sequencing

pipeline. Overall, piRNAs distribute randomly across the entire macronuclear chromosome with a slight depletion in subtelomeric regions (Figures 4 and S5).

Figure S5. Genome-wide mapping of piRNAs suggest that the whole somatic genome produces piRNAs, related to Figure 4 (panels A, D, G, J, M). Metagene analysis (N = 10,497) suggests that piRNAs distribute across entire macronuclear chromosomes at 12, 23, and 30 hr post-mixing with a small depletion in subtelomeric regions. In (J), 27 nt 5'-U reads from all four time points (12, 19, 23, and 30 hr) were combined. In (M), combined 27 nt 5'-U reads were mapped against 54 single-gene chromosomes that have no evidence of expression across the *Oxytricha* sexual cycle or in vegetative cells. For each mapping, the 13 nt position of a piRNA read is normalized against the length of the corresponding single-gene macronuclear chromosome, with red indicating the mRNA sense strand for protein-coding genes and blue indicating antisense strand. (B, E, H, K, N) Genome-wide nucleotide representation by piRNA reads. For each macronuclear chromosome surveyed, the nucleotide coverage is calculated as piRNA-covered sequence length divided by the contig length, and the distribution of nucleotide representation is plotted for all 10,497 contigs (B, E, H, K), or the 54 unexpressed genes (N). (C, F, I, L, O) Genome-wide nucleotide representation by piRNA reads mapping to either sense (red) or antisense (blue) strand of protein-coding genes, as calculated in (B, E, H, K, N). Only 27 nt 5'-U reads were used in the analyses. The difference in nucleotide representation likely reflects the sequencing depth of each time point (Table S2).

Figure S6. Small RNA injection leads to heritable retention of normally-deleted genomic regions, related to Figure 5. (A) MDS-IES junction sequences in synthetic RNA injection experiments. Pointers are direct repeat sequences present at the end of MDS(*n*) and the beginning of MDS(*n+1*). During genome rearrangement, one copy of the pointer sequence is retained in the new macronuclear genome, and one copy is deleted. (B) Single-cell PCR reveals efficient IES retention in individual F1 cells. Single 7-day exconjugants after injection were hand-isolated and used directly in PCR. Out of three sRNA injected cells, two demonstrate the exclusive presence of IES^+ nanochromosomes for Contig22226.0. The other cell contains both wild type (WT) and IES^+ versions. sDNA-injected and non-injected cells only produce a WT PCR product. (C) The protective effect of small RNA injection extends to multiple alleles. The top sequence shows the *actin I* MIC reference sequence, with annotations of MDS, IES, pointers, and the injected piRNA oligonucleotide immediately below it (T's shown instead of U's). Highlighted nucleotides are allelic differences that differ from the reference sequence. Bottom sequences are PCR clones from the sexual progeny of sRNA-injected cells. The C-to-T substitution highlighted in blue could be a PCR or sequencing error. (D, E) Substitutions present in injected synthetic piRNAs do not transfer to progeny DNA. The top sequence is the Contig22226.0 MIC reference sequence, with annotations of MDS, pointers, and injected sRNA below it (T's shown in place of U's). Shaded nucleotides D (panel D) and V (panel E) are non-C, and non-U respectively. Lower sequences are PCR clones from sRNA-injected cells. (F) Global piRNA mapping pattern to Contig22226.0 MAC and MIC contigs from a conjugation between WT cells (JRB310 and JRB510) or the Contig22226.0 $IES1^+$ strain backcrossed to wild-type strain JRB510. 26-28 nt 5'-U reads are plotted. The star indicates new IES-containing piRNAs produced in the backcrossed population.

Figure S1



B

```
  1 MESALTPIFP KRPVRGPKSQ GQELKLFSNY FSLEFDSPSI QGVNKYTCKF
 51 EPEVPDNSRK VRTLILRTVK DKIKESLDFF IEWGSCIYSL KKVQEVPKQE
101 TEHDGVKYAI TIEWVQLMEP TDKDHMNFLK IFFNSLMRGL RFETIGRKSF
151 NTAKAHSLDA HKIKVWPGFD ARLIMKETGV LLNIDVCFKV VRQDTVLEFI
201 NDLRSKCEQK NLDSQEEIAT ALKGTTVVTK YNQRTYKVDR VEFSMSPETT
251 FDKSGTQVSY KDYYKTRYNE NVSDPNQPLL INKDRKTGNE IALIPELCQV
301 TGLTDSMRAD FRLMKDLAEI VHTNADRKVS ECKNLLEIFN TNPKCLEKQK
351 LWHLKFSENP QALKGFKYKA GNMVMGAKGS GERNTFDIES CQREIDRKIQ
401 DKMFEQPALK TWGIFHGERD APICKQFTTT MEQVLQQFGF ESAQPQTFVL
451 KGNPMNAQTW IKELKSKLNN NVQACVLLLP GQKGKTSLYD DIKRFLLAEF
501 PVPSQVVLCN TISKGKNLRS IVNKILIQIT AKIGGIPWTV DALPFMEKPT
551 MICGMDVFHS TSLGKKSVLA LTASINTSAT KYFSSCVVQG ELGDEAANSL
601 QTGMSKALEA FKKANGSYPQ QMIFYRDGVG EGQVNGICVP EVEQIKAAIA
651 GLGIDVKFMY INVSKRINTR IFGGDVGSFK NPMPGTVIDA SITDSGIYEF
701 FLVSTAAKQG LSSPVRYSVI FDTIGESPDR IELLTYKLCY TYYNVSGSIK
751 EPSAIRYAHR LAALIGERGG KGKEPPQPHQ DFETKEPTLY YI
```

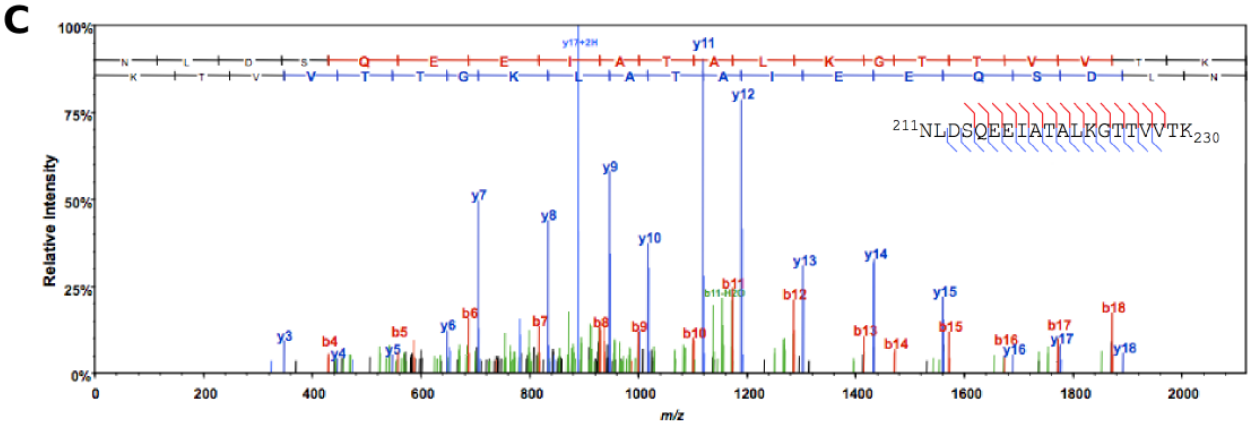


Figure S2

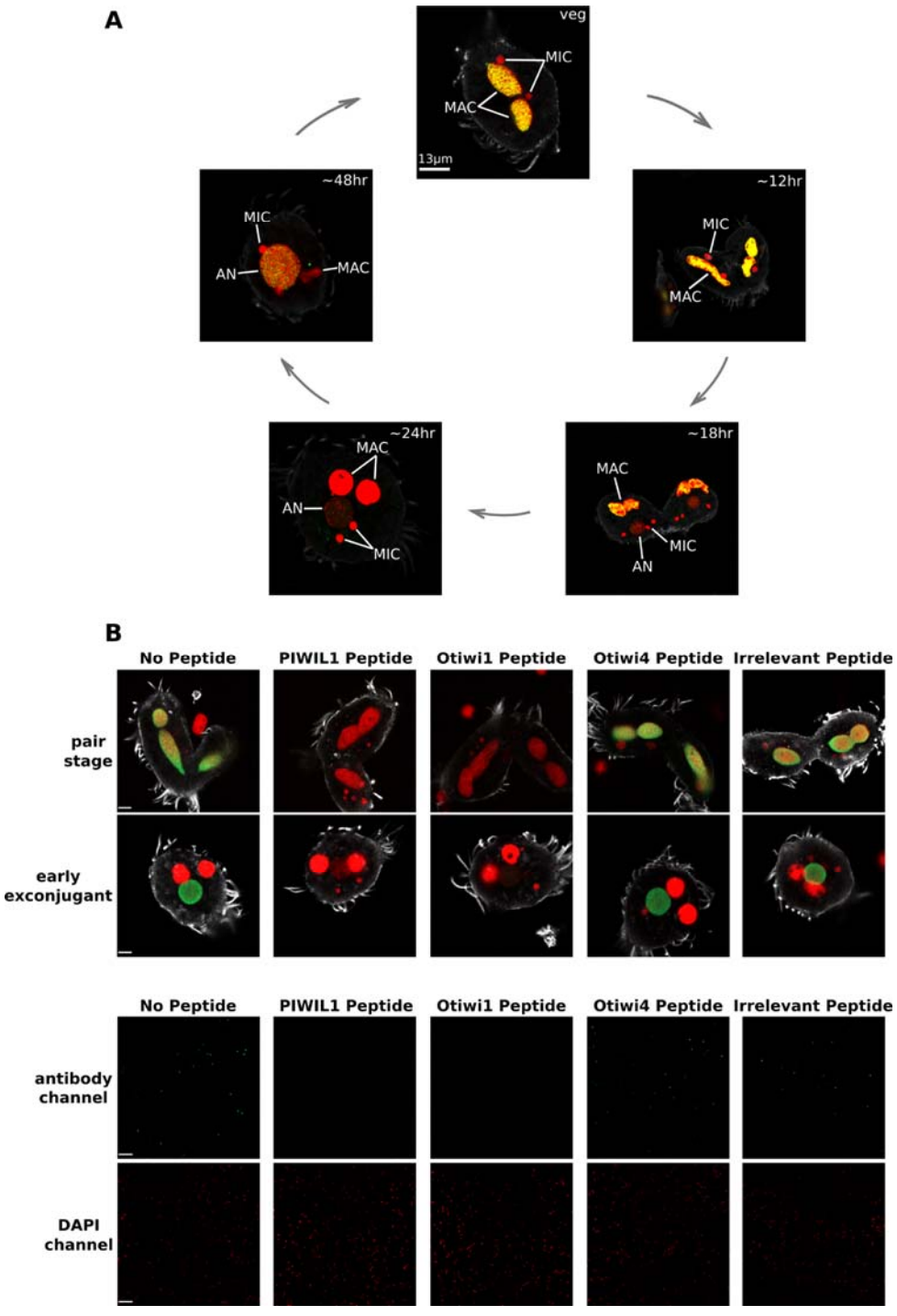

A
veg
MIC
MAC
13μm
~12hr
MIC
MAC
~18hr
MAC
MIC
AN
~24hr
MAC
AN
MIC
~48hr
MIC
AN
MAC
B
No Peptide
PIWIL1 Peptide
Otiwi1 Peptide
Otiwi4 Peptide
Irrelevant Peptide
pair stage
early exconjugant
antibody channel
DAPI channel

Figure S3

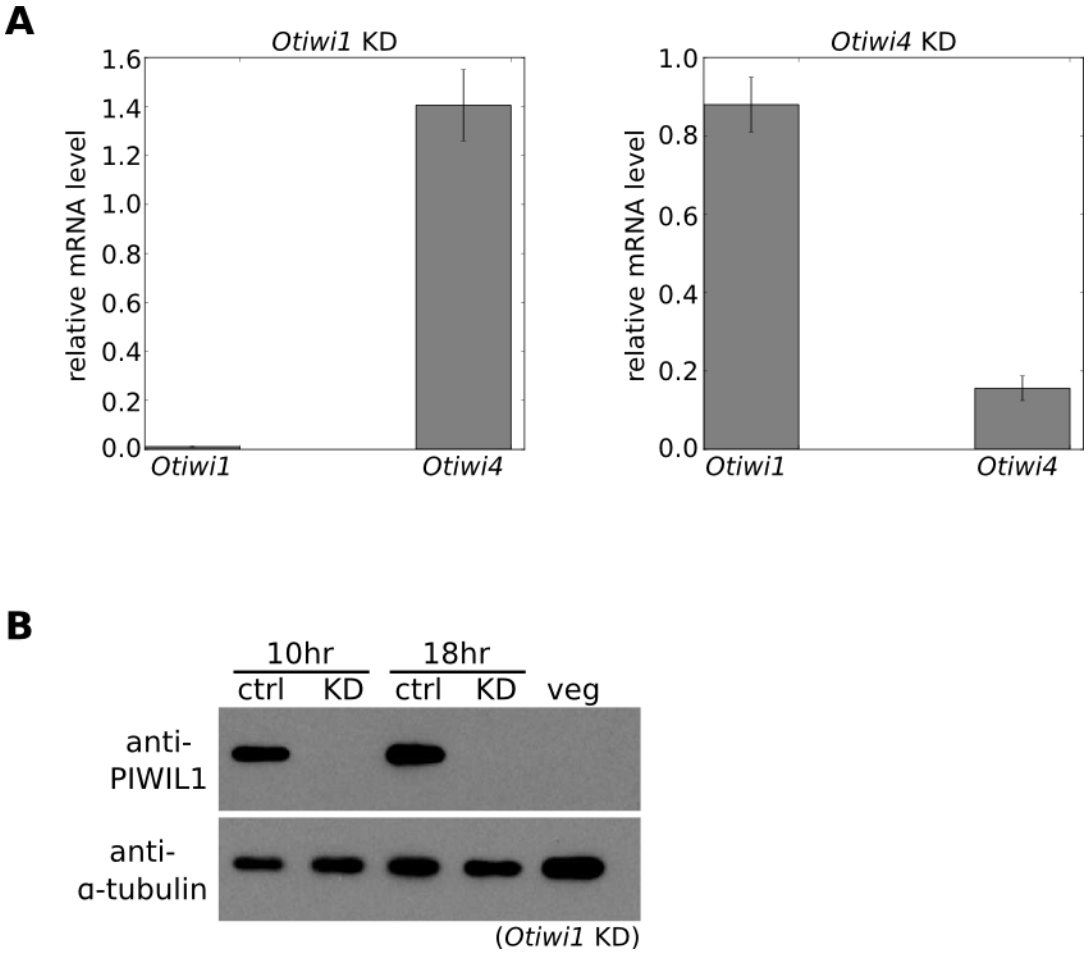

A
Otiwi1 KD
relative mRNA level
1.6
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
Otiwi1
Otiwi4
Otiwi4 KD
relative mRNA level
1.0
0.8
0.6
0.4
0.2
0.0
Otiwi1
Otiwi4
B
10hr
18hr
ctrl
KD
ctrl
KD
veg
anti-
PIWIL1
anti-
α-tubulin
(Otiwi1 KD)

Figure S4

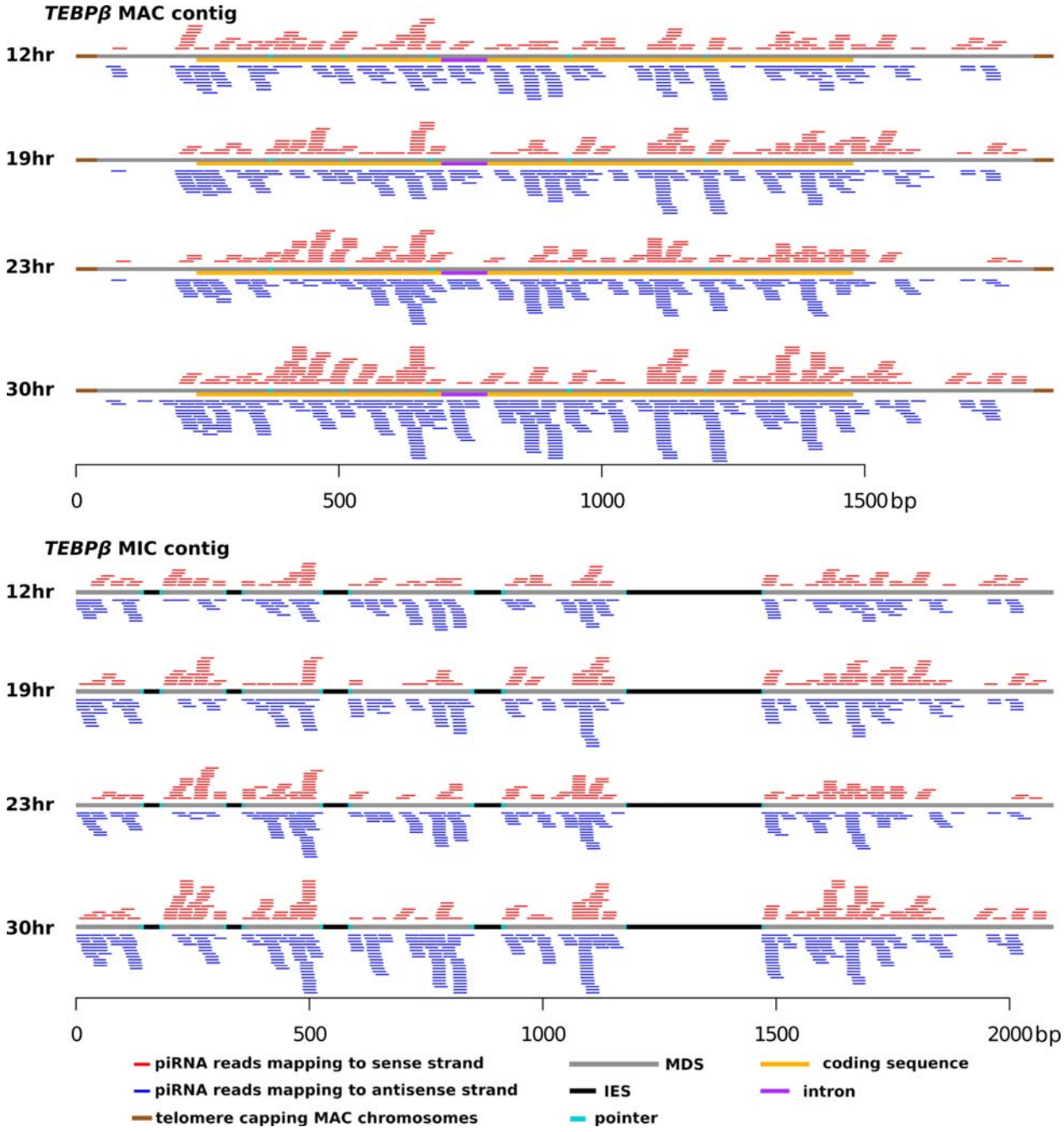

TEBPβ MAC contig
12hr
19hr
23hr
30hr
0
500
1000
1500 bp
TEBPβ MIC contig
12hr
19hr
23hr
30hr
0
500
1000
1500
2000 bp
piRNA reads mapping to sense strand
piRNA reads mapping to antisense strand
telomere capping MAC chromosomes
MDS
IES
pointer
coding sequence
intron

Figure S5

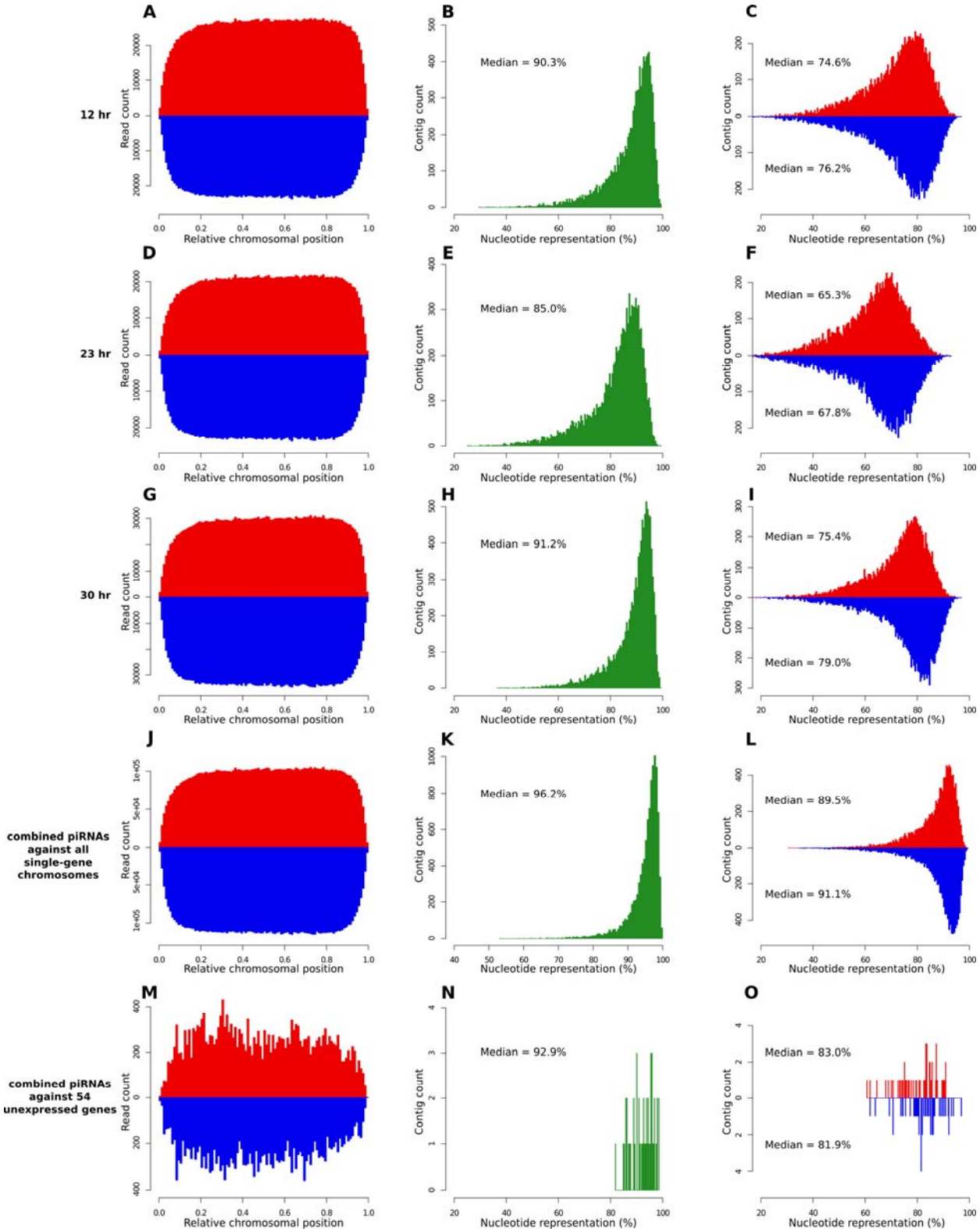
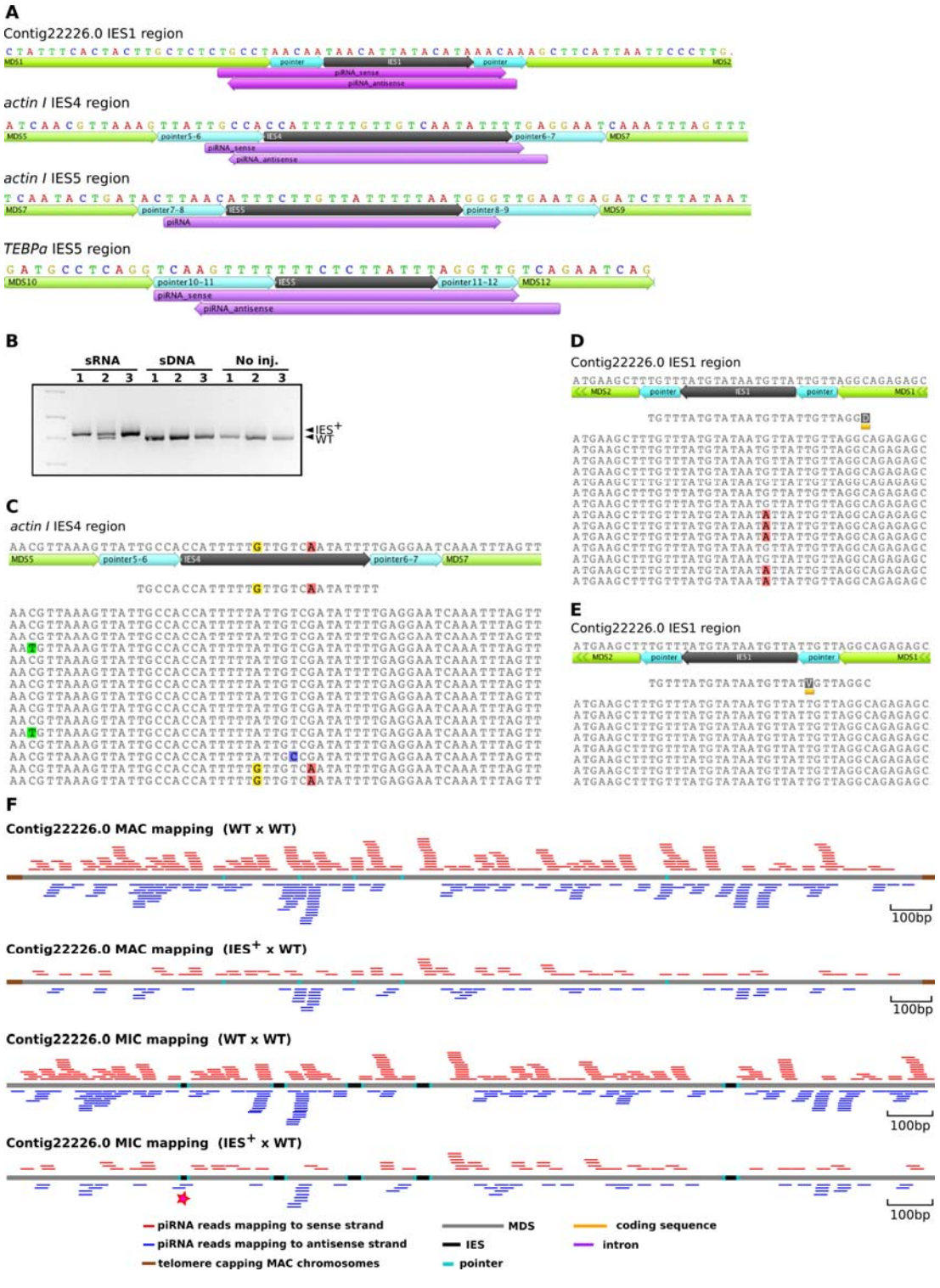

A
B
C
12 hr
Median = 90.3%
Median = 74.6%
Median = 76.2%
D
E
F
23 hr
Median = 85.0%
Median = 65.3%
Median = 67.8%
G
H
I
30 hr
Median = 91.2%
Median = 75.4%
Median = 79.0%
J
K
L
combined piRNAs against all single-gene chromosomes
Median = 96.2%
Median = 89.5%
Median = 91.1%
M
N
O
combined piRNAs against 54 unexpressed genes
Median = 92.9%
Median = 83.0%
Median = 81.9%
Read count
Contig count
Relative chromosomal position
Nucleotide representation (%)

Figure S6


A
Contig22226.0 IES1 region
actin I IES4 region
actin I IES5 region
TEBPa IES5 region
B
sRNA
sDNA
No inj.
1 2 3 1 2 3 1 2 3
IES+
WT
C
actin I IES4 region
D
Contig22226.0 IES1 region
E
Contig22226.0 IES1 region
F
Contig22226.0 MAC mapping (WT x WT)
Contig22226.0 MAC mapping (IES+ x WT)
Contig22226.0 MIC mapping (WT x WT)
Contig22226.0 MIC mapping (IES+ x WT)
100bp
piRNA reads mapping to sense strand
piRNA reads mapping to antisense strand
telomere capping MAC chromosomes
MDS
IES
pointer
coding sequence
intron

Supplemental tables

Table S1. Anti-PIWIL1 antibody primarily recognizes Otiwi1 during early conjugation, related to Figure 1.

		Number of unique peptides matched to MS/MS (within 0.7% peptide FDR)				Rough relative quantitative value (from spectra counting) of significant protein hits (within 0.7% peptide FDR)				Blastp hit (e value < 2e-12)	Pfam domain (e-value < 4e-10)
Identified Proteins	Molecular Weight	12hr	19hr	23hr	30hr	12hr	19hr	23hr	30hr		
Otwi1	89 kDa	62	86	97	69	134	139	138	137		
Otwi4	93 kDa	3	1	5	4	3	1	3	2		
Otiwi3	79 kDa	3	3	0	2	2	1	0	1		
Otiwi8	82 kDa	2	0	0	0	2	0	0	0		
Contig15376.0.g47.t1_0	91 kDa	2	3	9	5	85	141	71	47	nuclear cap binding protein	MIF4G;MIF4G_like_2
Contig22168.0.g50.t1_1	81 kDa	2	0	3	6	56	0	16	55	heat shock protein 90	HATPase_c, HSP90
Contig19613.0.g12.t1_0	134 kDa	0	0	3	2	0	0	16	16	transcription elongation factor SPT5-like	Spt5-NGN
Contig19200.0.g86.t1_1	141 kDa	0	0	4	2	0	0	22	16	RNA polII second largest subunit	RNA_pol_Rpb2_1~7
Contig9417.0.g16.t1_0	132 kDa	0	0	3	1	0	0	16	8	tRNA splicing endonuclease positive effector-related protein; helicase sen1	AAA_12

Search table of mass spectrometry raw data for the ~90kD protein band against all Otiwi proteins and a predicted *Oxytricha* proteome (Swart et al., in revision) at 12, 19, 23, and 30 hr post-mixing.

Table S2. Number of reads after each step of sRNA data processing, related to Figure 4 and Figure S6.

	Raw reads	Adapter trimmed	Non-redundant	Telomere filtered
12 hr IP	59,613,040	38,422,470	16,136,476	16,134,428
19 hr IP	54,845,092	35,431,903	14,747,393	14,746,015
23 hr IP	24,538,653	22,821,304	12,257,923	12,255,581
30 hr IP	43,052,455	40,128,491	18,783,423	18,778,921
20 hr total	25,800,567	24,120,916	3,427,458	3,426,947
19 hr backcross total	7,567,730	7,324,632	4,853,791	4,853,345

Extended Experimental Procedures

Crosslinking RNA co-immunoprecipitation

To extract Otiwi1-bound RNA, ~5 x 10^6 cells (~200 µl pellet) were collected at 12, 19, 23, and 30 hrs post mixing. Crosslinking was carried out in 1% formaldehyde for 10 minutes at room temperature and then quenched by 125 mM glycine, final concentration. Cells were washed once with TBS (50 mM Tris-Cl pH 7.5, 140 mM NaCl, 2 mM EDTA) and once with IP buffer (TBS supplemented with 1% NP-40, 0.1% deoxycholate, complete protease inhibitor cocktail tablets from Roche, and SUPERase•In from Ambion). Cells were collected by centrifugation at 1,000g for 5 min, and pellets were resuspended in IP buffer and sonicated (Branson Sonifier 450, output 3, duty cycle 30%, 10 pulses repeated 10 times). Lysates were cleared by centrifugation at 12,000 g for 15 minutes, and supernatant was collected. 7.5 µg Piwi antibody (ab12337, Abcam) or normal rabbit IgG (12-370, Millipore) was added to lysates and incubated for 3 hr at 4 °C. Dynabeads coated with protein G (Invitrogen) were then added and the incubation was continued for overnight at 4 °C. The beads were washed at room temperature for 5 minutes with each of the following buffers: SDS wash (50 mM Tris-Cl pH 7.5, 140 mM NaCl, 2 mM EDTA, 0.025% SDS), high salt wash (50 mM Tris-Cl pH 7.5, 500 mM NaCl, 2 mM EDTA), LiCl wash (20 mM Tris-Cl pH 7.5, 250 mM LiCl, 2 mM EDTA), TE wash (20 mM Tris-Cl pH 7.5, EDTA 2 mM). Immunoprecipitated complex was eluted from the beads in 0.1 M $NaHCO_3$ and 1% SDS supplemented with SUPERase•In (Ambion) at 65 °C for 30 min. Proteinase K (Roche, 1µg/µl) was immediately added to digest protein. Eluted complex was reverse crosslinked at 65 °C for 2.5 hr, after the addition of NaCl (200 mM final concentration) and EDTA (10mM). RNA was phenol extracted and ethanol precipitated.

Mass Spectrometry-based proteomic analysis of proteins immunoprecipitated with the anti-Otiwi1 antibody

Crosslinking IP was performed as described above up to the Proteinase K digestion step. The eluted complex in 1% SDS and 0.1M $NaHCO_3$ was incubated at 65 °C for 7 hr for reverse crosslinking. The reverse-crosslinked eluates were directly separated by 4-12% NuPAGE® Bis-Tris Gel (Invitrogen) and stained with GelCode Blue Safe Protein Stain (Thermo Scientific) following manufacturer's instructions. The ~ 90kD bands from all four time points and ~ 50kD band from 23hr IP were cut out and subjected to in-gel cysteine reduction and alkylation and

trypsin digestion. Peptides were eluted and subjected to LC-MS/MS on a ThermoFisher LTQ-Orbitrap XL MS instrument platform equipped with an Eksigent nano-flow uPLC system and an Advion NanoMate ion source robot at the Princeton University Mass Spectrometry Center. Peak lists from the resulting MS/MS data were generated using ProteomeDiscoverer software (ThermoFisher) and subjected to database search against a predicted *Oxytricha* proteome derived from the macronuclear genome annotation (Swart et al., in revision), including a subdatabase containing only Otiwi proteins confirmed in this study, using the Mascot Search engine (Matrix Science), allowing for a single missed cleavage, cysteine carbamidomethylation, methionine oxidation, and N-terminal protein acetylation. Peptide and protein assignments were collated, consolidated using the principles of protein parsimony, restricted by a score cutoff yielding a peptide false-discovery rate of <1%, and subjected to spectral counting quantitation using the Scaffold software (Proteome Software).

Agarose Northern analysis

Total RNA was separated on a 1% denaturing agarose gel with 2.2 M formaldehyde in 1× MOPS buffer. RNA was transferred to Hybond XL membrane (Amersham) overnight in 10 × SSC using a Nytran TurboBlotter (Schleicher & Schuell). DNA probes were generated by random priming (Invitrogen) of PCR products corresponding to respective coding regions. After overnight hybridyzation in Church buffer at 60 °C, the membrane was washed in 2X SSC with 0.1% SDS for 30 min at 55 °C.

Western analysis

Cell pellets were directly lysed in Laemmli buffer. The lysates were separate by 10% SDS-PAGE, and protein transferred to Immun-Blot PVDF Membranes (Bio-Rad) with a Trans-Blot SD Semi-Dry Transfer Cell (Bio-Rad) following manufacturer's instructions. The membrane was blocked with 5% non-fat milk at room temperature for 30 min, incubated with Otiwi1 antibody (1:1000 dilution in 5% non-fat milk) overnight at 4 °C. The membrane was washed with TBST for 15 min four times, and incubated with mouse-anti-rabbit secondary antibody (1:10,000 dilution, Jackson Immunoresearch Laboratories) for 1hr, washed again with TBST for 15 min four times. The signal was detected with SuperSignal West Pico Chemiluminescent Substrate (Thermo Scientific) following manufacturer's instructions. After the detection of Otiwi1, the membrane was stripped with Restore Western Blot Stripping Buffer (Thermo Scientific) and re-probed with the anti-α-tubulin antibody and goat-anti-mouse secondary antibody (Jackson Immunoresearch Laboratories) as described above.

Immunostaining

Oxytricha cells were fixed in 4% paraformaldehyde (Electron Microscopy Sciences) for 10 min at room temperature, washed with PBS twice and settled onto 2% polylysine-coated slides overnight at 4 °C. Fixed cells were then permeabilized with 0.5% triton X-100 in PBS for 20 minutes at room temperature, and incubated in 0.1 N HCl for 5 minutes. Image-iT signal enhancer (Invitrogen) was used for blocking for 30 min. Cells were incubated with primary antibodies diluted in 5% BSA in PBS for 1 hr at room temperature, washed in PBS for 20 min, and then incubated with secondary antibodies for 1.5 hr at 37°C. DNA was stained with DAPI (1:2000 diluted in PBS). Samples were mounted in Aqua-Poly/Mount (Polysciences). The following primary antibodies and concentrations were used: anti-PIWIL1 (ab12337, Abcam, 1:500), anti-alpha tubulin (ab7291, Abcam, 1:1000). Secondary antibodies are mouse-anti-rabbit Alexa 488 and goat-anti-mouse Alexa 568 (Invitrogen). Laser scanning confocal images are

taken with Zeiss LSM510 at the Microscopy Facility in Molecular Biology, Princeton University.

Peptide competition assay
The four peptides used are human PIWIL1 peptide (15aa C-terminal peptide, from Abcam ab13827); Otiwi1 C-terminal peptide: PHQDFETKEPTLYYI (LifeTein); Otiwi4 C-terminal peptide: QVHMKFEDSSGLYFI (LifeTein); irrelevant peptide: LSKINFRPLPHDPKVHFEK (Thermo Scientific). Peptides were incubated at 4 °C overnight with the antibody at a 2000:1 molar ratio (8.8 μM for peptide and 4.4 nM for antibody, respectively), in immunostaining blocking buffer. The pre-incubated antibody solutions were then used directly for immunostaining as described above, with a 2-fold dilution when applied to cells.

Design of antisense oligonucleotides for knockdown
mRNA local structures were predicted using mFold with default parameters, and the unpaired regions (loops, bulges, and mismatches) supported by multiple low energy folding were chosen for antisense oligonucleotide end annealing. In the knockdown experiments, we mixed eight different oligonucleotides with phosphorothioate linkages in their backbones (Integrated DNA Technologies, IDT).

RT-qPCR verification of knockdown
Extracted RNA was DNase treated with TURBO DNA-free kit (Ambion). First strand synthesis of cDNA was carried out using SuperScript III (Invitrogen) with oligo dT primers. The 7900HT Fast Real-Time PCR System and SYBR green master mix (Applied Biosystems) were used for qPCR analysis with the default cycling program. Primer specificity was confirmed by melt-curve analysis. ΔΔCt method was used to calculate the fold-change of cDNA levels.

Small RNA sequencing data processing and analysis
The raw sequencing data for Otiwi1-associated sRNAs were first barcode-split by using Barcode Splitter (version1.0.0) on Princeton Galaxy web server (Blankenberg et al., 2010; Giardine et al., 2005; Goecks et al., 2010). Then 3' adapter trimming requires a perfect 21 nt match to "TGGAATTCTCGGGTGCCAAGG". The 20 hr total sRNA sequencing reads were trimmed by requiring a perfect match to "ATCTCGTATGCCGTC". 19 hr sRNAs from Contig22226.0 IES1^{+} strain backcrossed to wild-type strain were retrieved from pooled sequencing by the perfect match to the barcode sequences "TCACGATCAGATCTC" and "TCACTAGCTTATCTC", and adapter-trimmed requiring a perfect match to "TGGAATTCTCGGGTG". Reads that contain "N" or were shorter than 18bp were discarded. The remaining reads were collapsed using FASTX Tool kit Collapser (http://hannonlab.cshl.edu/fastx_toolkit/index.html), and those that contain telomeric sequences "CCCCAAAACCCC" or "GGGGTTTTGGGG" were removed. The non-redundant telomere-filtered datasets were used for all subsequent analyses. The sequencing and primary processing statistics are summarized in Table S2.

For mapping, a collection of genomic sequences was merged to produce a "master" database, which includes the *O. trifallax* strain JRB310 MAC genome assembly (Swart *et al*. in revision), *O. trifallax* rRNA MAC contigs, TBE1 transposon insertion sequences of the TBE1tri1 locus (GenBank: L39906.1), the *O. trifallax* mitochondria genome (Swart et al., 2012), and the *Chlamydomonas reinhardtti* genome (Merchant et al., 2007). Mapping of the reads against the "master" database was carried out with SHRiMP version 2.1.0 (David et al., 2011) using the

default parameters except for "-U" for ungapped mapping, "-R" for appending reads in the output, and "--strata" for outputting the best hits only. After SHRiMP mapping, a customized program was used to remove mappings with more than two mismatches. We allow two mismatches to accommodate allelic differences between the genomes of mating-compatible *O. trifallax* strains JRB310 and JRB510. Then a hierarchical filtering was used to partition reads to the target genome, in the order of rRNA, mitochrondrial genome, *Chlamydomonas* genome, TBE1, and MAC genome. This means that if a read maps to rRNA, then it is "assigned" to rRNA even if it also maps to other genomes or to loci with the same mapping score. After this hierarchical filtering, we removed redundant mapping, assigning piRNA mapping randomly in the instance where a read maps to multiple loci with the same score. We then calculated the fraction of 27 nt 5'-U reads mapping to each genome/locus (Figure 4C). Since we observed that most rRNA reads are also 27 nt and contain 5'-U, suggesting that they are likely authentic piRNAs rather than degradation products, we ultimately returned the rRNA fraction to MAC-mapping category.

Customized R programs were used to visualize piRNAs (27 nt with 5'-U only) mapping to specific loci (Figures 4D and S4). To characterize genome-wide piRNA mapping positions (Figures 4E, F, G and S5), we analyzed piRNAs (27 nt with 5'-U only) mapping to a subset of the MAC genome that consists of 10,497 two-telomere-containing, single-gene chromosomes (the 54 genes unexpressed in our time points are a subset of these 10,497 genes) oriented to the sense strand of the predicted mRNAs. We also truncated the telomeric sequences from each contig, such that the chromosome sequences start and end with the subtelomeric regions. Distributions of absolute number of reads mapped to each bin of relative position on a chromosome were plotted in Figure 4E and Figures S5A, D, G, J, M. For each macronuclear chromosome surveyed, the nucleotide coverage is calculated as piRNA-covered sequence length divided by the contig length, and the distribution of nucleotide representation is plotted for all 10,497 contigs or the 54 unexpressed genes (Figure 4F, Figures S5B, E, H, K, N), which were split to sense and antisense piRNA representation (Figure 4G, Figures S5C, F, I, L, O).

The master genome sequence, single-gene contig sequences, as well as all mapping outputs are available at http://trifallax.princeton.edu/cms/databases/data-files/otiwi1_manuscript_relevant.

Oligonucleotides (5' - 3')

Primers for RT-PCR of *Otiwi* genes

Otiwi1_ORF_F	ATGGAATCCGCTTTGACACCAATC
Otiwi1_ORF_R	TCAAATGTAGTAGAGGGTTGGCTCC
Otiwi2_ORF_F	ATGGAACTTTAATTGGCAAAGAGACCAAG
Otiwi2_ORF_R	TCAAATGAAATAAAGCCCTTGAATGTTCTTAGAG
Otiwi3_ORF_F	ATGAGTAAAGCTTCACCAGAATCAATTAATCAAG
Otiwi3_ORF_R	TCAGATATAGTAGAGACCTTTGATTTTCTCATAATGATTA
Otiwi4_ORF_F	ATGGAAAGTAAGATATAAACTTAACCAAGAAAAAATGAGC
Otiwi4_ORF_R	TCAAATGAAATAAAGGCCACTTGAGTCCT
Otiwi5_ORF_F	ATGTCGCATTTTAACTATAACAATCCAAACAATCAAC
Otiwi5_ORF_R	TCATAAGAAATGTAGTGATTGACTTAACTTAGGGTTC
Otiwi6_ORF_F	ATGGATTACAGAATGGGCGGAGGA
Otiwi6_ORF_R	TCACAAGAAAAACGGTAGTCTGTGCATATC
Otiwi7_ORF_F	ATGGAACGAAGAAACAACAACCAAGG
Otiwi7_ORF_R	TCAAAGGTAGTGGAAAGATTCGTAGACTTTC
Otiwi8_ORF_F	ATGGAAATTGACCTCAACATTAAGTACAGAGC
Otiwi8_ORF_R	TCATAAGAAATGCAAGTTGAATGCAAGTTTGG
Otiwi9_ORF_F	ATGGAATCATACTCAAAACCAAACAATTAAAGAAAC
Otiwi9_ORF_R	TCATAAGAAATGCAGAGACTCTTGCATG
Otiwi10_ORF_F	ATGGACCGCAGAAAGAATCCAACT
Otiwi10_ORF_R	TCATGTAATTAGCATAGTGGTTTAGCACCG
Otiwi11_ORF_F	ATGTTCAAAGACCGCAAGACAGACC
Otiwi11_ORF_R	TCACAAATAATGCAATCCATGTATATTTTTATCAAAGTG
Otiwi12_ORF_F	ATGGAACGAAGCTTTAATAGCAGTAGTATTCTTGG
Otiwi12_ORF_R	TCACAAATAATGCAAATTAAAAGCAAGTTTTTCATTTGG
Otiwi13_ORF_F	ATGAATTCAACAGTGAGAGAGGAACCAG
Otiwi13_ORF_R	TCATAAGAAATGCAATGTAAAATCAAGATTTTCGTTTGG

Primers for RT-PCR of *actin I*

Actin I_M3F	**AGGACGCTCCAAGAGCTGTATTCC**
Actin I_M6R	**CCACATGCTGGCGAAGGTTGA**

Antisense DNA oligonucleotides for knockdown (* indicate phosphorothioate linkage)

Otiwi1

C*T*T*T*G*A*A*G*C*A*A*A*C*A*T*C*A*A*T*G*T*T*C
C*T*G*G*T*C*T*T*G*T*A*G*T*A*G*T*C*T*T*T*G
G*A*G*C*T*T*G*A*G*G*A*T*T*T*T*C*C*G*A*G*A
G*C*A*T*A*C*T*T*G*A*C*T*C*C*A*T*C*A*T*G*T*T*C
A*T*T*C*C*A*G*C*A*C*T*G*T*A*T*C*T*T*G*G*C
A*T*T*C*C*A*G*C*A*C*T*G*T*A*T*C*T*T*G*G*C*G*C*A*C
C*A*T*G*T*G*G*T*C*C*T*T*G*T*C*T*G*T*T*T*G*G
C*T*T*A*G*C*T*G*T*G*T*T*G*A*A*A*C*T*C*T*T*C*C

Otiwi4
T*G*T*G*A*C*A*A*G*G*C*T*C*G*A*T*A*T*T*C*A*A
G*T*C*C*G*T*A*T*C*T*T*T*G*T*T*G*A*G*C*A*A
T*A*C*T*C*A*A*G*T*G*C*A*G*A*T*T*C*G*T*A*T*C*T
T*T*C*T*G*T*C*G*A*G*G*T*C*T*C*T*T*T*C*C*A*T*T
C*C*C*A*T*T*T*C*T*G*A*A*T*A*G*C*T*G*G*T*T
A*G*T*A*T*G*A*G*G*A*C*T*A*C*G*C*A*C*T*A*G
C*H*A*C*A*A*A*A*C*T*T*G*C*C*A*T*C*T*G*T*T*T*G*G*C
G*G*T*T*G*T*A*T*C*T*C*A*A*A*G*T*A*G*T*T*A*G*C*G*A*A

Control oligonucleotide for knockdown (* indicate phosphorothioate linkage)

A*A*A*G*A*T*G*A*C*G*G*G*A*A*C*T*A*C*A*A*G*A*C*A

Primers used for qPCR verification of *Otiwi1* and *Otiwi4* knockdown

Otiwi1_F	CCTGAGATTTGAGACCATCGG
Otiwi1_R	AGTCTAGCATCAAATCCAGGC
Otiwi4_F	CCCATTCCTGGAGTTGTTGTGAAC
Otiwi4_R	ATCTTGTAGGCTTCACCATGCCCT
mito_rRNA_F	CATATCCTGGTTGTGAATAATCTTCCAAGGG
mito_rRNA_R	GATAGGGACCGAACTGTCTCACG

Synthetic RNA oligonucleotides used in IES-retention injection experiments:

Contig22226.0_IES1_as
/5Phos/rUrGrUrUrUrArUrGrUrArUrArArUrGrUrUrArUrUrGrUrUrArGrGrC
Contig22226.0_IES1_sense
/5Phos/rUrGrCrCrUrArArCrArArUrArArCrArUrUrArUrArCrArUrArArArC
Contig22226.0_IES1_DNA
TGTTTATGTATAATGTTATTGTTAGGC
denegerate_piRNA_UW (W = A or U)
/5Phos/rUrWrWrWrWrWrWrWrWrWrWrWrWrWrWrWrWrWrWrWrWrWrW
Contig22226.0_IES1_as2 (for generating the stable line in the backcross and sequencing study)
/5Phos/rUrArUrGrUrArUrArArUrGrUrUrArUrUrGrUrUrArGrGrCrArGrArG
Actin_I_IES4_sense
/5Phos/rUrGrCrCrArCrCrArUrUrUrUrUrGrUrUrGrUrCrArArUrArUrUrU
Actin_I_IES4_as
/5Phos/rUrCrArArArArUrArUrUrGrArCrArArCrArArArArArUrGrGrUrGrG
Actin_I_IES4_DNA
TGCCACCATTTTTGTTGTCAATATTTT
Actin_I_IES5_sense
/5Phos/rUrUrArArCrArUrUrUrCrUrUrArUrUrArUrUrUrUrUrArArUrGrGrG
TEBPα_IES5_sense
/5Phos/rUrCrArArGrUrUrUrUrUrUrUrCrUrCrUrUrArUrUrUrArGrGrUrUrG
TEBPα_IES5_as
/5Phos/rUrGrArCrArArCrCrUrArArArUrArArGrArGrArArArArArArCrU

Primers for detecting IES$^+$ DNA post synthetic oligonucleotide injection:

For Contig22226.0 IES1 detection:

Contig22226.0_M1_F	TGCTATTTCACTACTTGCTCTCTGCC
Contig22226.0_M4_R	TGACGAAGATCAAGAGCATTCTCAAATCTAG

For *actin I* IES4 detection:

Actin_I_M4_F	GTACGAAGGTATCGGTGAGAGACTTC
Actin_I_M9_R	GCACACATTATAAAGATCTCATTCAACCC

For *actin I* IES5 detection:

Actin_I_M4_F	GTACGAAGGTATCGGTGAGAGACTTC
Actin_I_i5M9_R	

CTAAGACTGTGCACACATTATAAAGATCTCATTCAACCCATTAAAAATAA

For TEBPα IES5 detection:

TEBPα_M9_F	CTTCACCCAATACTCAGTTATCTCC
TEBPα_M12_R	GTAGCTGATCTGATTCTGACAACC

References

Blankenberg, D., Von Kuster, G., Coraor, N., Ananda, G., Lazarus, R., Mangan, M., Nekrutenko, A., and Taylor, J. (2010). Galaxy: a web-based genome analysis tool for experimentalists. Curr Protoc Mol Biol *Chapter 19*, Unit 19 10 11-21.

David, M., Dzamba, M., Lister, D., Ilie, L., and Brudno, M. (2011). SHRiMP2: sensitive yet practical SHort Read Mapping. Bioinformatics *27*, 1011-1012.

Giardine, B., Riemer, C., Hardison, R.C., Burhans, R., Elnitski, L., Shah, P., Zhang, Y., Blankenberg, D., Albert, I., Taylor, J.*, et al.* (2005). Galaxy: a platform for interactive large-scale genome analysis. Genome Res *15*, 1451-1455.

Goecks, J., Nekrutenko, A., and Taylor, J. (2010). Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. Genome Biol *11*, R86.

Merchant, S.S., Prochnik, S.E., Vallon, O., Harris, E.H., Karpowicz, S.J., Witman, G.B., Terry, A., Salamov, A., Fritz-Laylin, L.K., Marechal-Drouard, L.*, et al.* (2007). The Chlamydomonas genome reveals the evolution of key animal and plant functions. Science *318*, 245-250.

Swart, E.C., Nowacki, M., Shum, J., Stiles, H., Higgins, B.P., Doak, T.G., Schotanus, K., Magrini, V.J., Minx, P., Mardis, E.R.*, et al.* (2012). The Oxytricha trifallax mitochondrial genome. Genome Biol Evol *4*, 136-154.